

ARNOLD, WHITE & DURKEE

A PROFESSIONAL CORPORATION

Attorneys at Law

POST OFFICE BOX 4433

HOUSTON, TX 77210-4433

750 BERING DRIVE, SUITE 400

HOUSTON, TX 77057-2198

TELEPHONE (713) 787-1400

FACSIMILE (713) 789-2679

1900 ONE AMERICAN CENTER
600 CONGRESS AVENUE
AUSTIN, TX 78701-3248
TELEPHONE (512) 418-3000
FACSIMILE (512) 474-7577

155 LINFIELD DRIVE
MENLO PARK, CA 94025-3741
TELEPHONE (415) 614-4500
FACSIMILE (415) 614-4599

4850 FIRST BANK PLACE
601 SECOND AVE SOUTH
MINNEAPOLIS, MN 55402-4320
TELEPHONE (612) 321-2800
FACSIMILE (612) 321-9600

800 QUAKER TOWER
321 NORTH CLARK STREET
CHICAGO, IL 60610-4714
TELEPHONE (312) 744-0090
FACSIMILE (312) 755-4489

2001 JEFFERSON DAVIS HIGHWAY
SUITE 401
ARLINGTON, VA 22202-3604
TELEPHONE (703) 415-1720
FACSIMILE (703) 415-1728

October 15, 1997

FILE: IOWA:012/FUS

WRITER'S DIRECT DIAL:
512/418-3050

EXPRESS MAIL MAILING LABEL
NUMBER EM465544775US
DATE OF DEPOSIT October 15, 1997
I hereby certify that this paper or fee is being deposited with the United States Postal Service "EXPRESS MAIL POST OFFICE TO ADDRESSEE" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to: Assistant Commissioner for Patents, Washington D.C. 20231.
<i>Sandra Cherry</i> Sandra Cherry

BOX PATENT APPLICATION

Assistant Commissioner for Patents
Washington, DC 20231

RE: *U.S. Patent Application Entitled: P-TEFb COMPOSITIONS, METHODS AND SCREENING ASSAYS - David H. Price*

Sir:

Transmitted herewith for filing is a 294-page patent specification including 67 claims and an abstract. Also included are Figures 1-8 on 8 sheets. The specification and drawings constitute the application of David Price for the captioned application.

Also transmitted herewith is a diskette containing the computer-readable form of those sequences in the specification, a Statement as Required Under 37 C.F.R. § 1.821(f), and a separate paper copy of the sequence listing.

Please note that this application is filed without an inventor's Declaration and Assignment, a Declaration Claiming Small Entity Status, a Power of Attorney, and filing fees. Pursuant to 37 C.F.R. § 1.53(b) and (d), the Applicant requests the Patent and Trademark Office to accept this application and accord a serial number and filing date as of the date this application

ARNOLD, WHITE & DURKEE

Assistant Commissioner for Patents
October 15, 1997
Page 2

is deposited with the U.S. Postal Service for Express Mail. Further, the Applicant requests that the NOTICE OF MISSING PARTS-FILING DATE GRANTED pursuant to 37 C.F.R. § 1.53(d) be sent to the undersigned Applicant's representative.

Please date stamp and return the enclosed postcard to evidence receipt of this application.

Respectfully submitted,



Shelley P.M. Fussey
Patent Agent
Reg. No. 39,458

SPMF:JAS:jhb:sac

Encl:

PATENT
IOWA:012

APPLICATION FOR UNITED STATES LETTERS PATENT

for

P-TEFb COMPOSITIONS, METHODS AND SCREENING ASSAYS

by

David H. Price

EXPRESS MAIL MAILING LABEL	
NUMBER	EM465544775US
DATE OF DEPOSIT	<u>10/15/97</u>
I hereby certify that this paper or fee is being deposited with the United States Postal Service "EXPRESS MAIL POST OFFICE TO ADDRESSEE" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to: Assistant Commissioner for Patents, Washington D.C. 20231.	
<u>Sandra Cherry</u>	
Sandra Cherry	

BACKGROUND OF THE INVENTION

The present application claims priority to co-pending provisional application Serial No. 60/029,760, filed October 25, 1996, the entire text and figures of which disclosure are specifically incorporated herein by reference without disclaimer. The U.S. Government owns rights in the present invention pursuant to grant number R01-GM35500 from the National Institutes of Health.

1. Field of the Invention

The present invention relates generally to the fields of cellular biochemistry and viral replication. More particularly, it concerns the discovery that a certain factor, P-TEFb, has a central role in transcription elongation control, that it phosphorylates RNA polymerase II and that it binds to the HIV protein, Tat. The invention provides human genes encoding the P-TEFb subunits, various other biological components, and methods relating to the control of transcription elongation that have particular utility in the identification of substances that inhibit viral replication.

2. Description of Related Art

The production of any functional eukaryotic mRNA requires efficient transcription elongation by RNA polymerase II. Eukaryotic gene expression is controlled in part during the elongation phase of transcription. Shortly after initiation, RNA polymerase II acquires the properties necessary to synthesize full length pre-mRNAs (Spencer and Groudine, 1990b; Kerppola and Kane, 1991; Wright, 1993; Bentley, 1995; Maldonado and Reinberg, 1995). As is frequently found in control processes, there is a negative control mechanism which is manifest as a blockage during early elongation.

Blocks in transcription, usually referred to as premature termination, have been observed during transcription in a number of systems, including mammalian genes such as α -tubulin (Middleton and Morgan, 1990; Hair and Morgan, 1993), and in viruses, such as adenovirus

(Kessler *et al.*, 1989), simian virus 40 (SV40) (Kessler *et al.*, 1991), minute virus of mice (Krauskopf *et al.*, 1991) and human immunodeficiency virus (HIV) (Laspia *et al.*, 1989). RNA polymerase II molecules are also found blocked, during elongation, near the promoter on many genes in *Drosophila melanogaster* (Rougvie and Lis, 1988; 1990).

5

Except for the involvement of the viral Tat protein in HIV gene expression (Marciniak and Sharp, 1991), little is known about the molecular mechanisms involved in elimination of this block. *In vivo*, HIV transcription is tightly controlled by the viral Tat protein. It is known that Tat acts as a potent transcriptional transactivator by binding to the transactivation response (TAR) region on the nascent RNA and interacting with cellular factors (Garcia and Gaynor, 1994; Jones and Peterlin, 1994). Still, many issues remain to be clarified concerning the interactions and functional regulation of transcriptional elongation and Tat, and further information is needed before effective anti-HIV strategies can be developed based upon intervention connected with Tat activity.

15

Concerning cellular genes, the transcription of the proto-oncogene, *c-myc*, is regulated by a block during elongation (Miller *et al.*, 1989; Spencer and Groudine, 1990a; Wright and Bishop, 1989). *C-myc* expression was studied in *Xenopus* oocytes and isolated HeLa nuclei. The block to elongation in *c-myc* occurs close to the promoter, termed "promoter-proximal pausing", and only short RNAs are produced (Krumm *et al.*, 1995; Meulia *et al.*, 1993; Strobl and Eick, 1992). A block to elongation at the end of the first exon regulates the levels of *c-fos* RNA in response to tumor promoters and intracellular calcium levels (Collart *et al.*, 1991; Mechti *et al.*, 1991). Other blocks to elongation occur in the transcription of the proto-oncogenes *c-myb* (Bender *et al.*, 1987; Reddy and Reddy, 1989) and *c-fms* (Yue *et al.*, 1993). The RNA levels for the adenosine deaminase genes (ADA) of humans and mice are at least partly controlled by a regulated block to elongation (Ramamurthy *et al.*, 1990; Chinsky *et al.*, 1989; Chen *et al.*, 1990; Chen *et al.*, 1991; Kash *et al.*, 1993). Thus this elongation control process has been implicated in the expression of many genes, yet the mechanism of control is not yet understood.

Studies in human, murine, *Drosophila* and *Xenopus* systems have demonstrated the existence of two classes of elongation complexes differing in their potential to produce full length mRNA sized transcripts. A model for the control of elongation has been described which is based, in part, on results obtained from a *Drosophila* *in vitro* transcription system (Kephart *et al.*, 1992; Marshall and Price, 1992) and is consistent with data obtained *in vitro* and *in vivo* from many studies.

Key features of the elongation control model are that all RNA polymerase II molecules that initiate from a promoter are destined to produce only short transcripts, in a process termed "abortive elongation". Abortive elongation is distinct from abortive initiation because the abortive transcripts are 10 to 20 times longer during abortive elongation and, presumably, the polymerase in the abortive elongation complexes must relocate the promoter after producing an abortive transcript to bring about reinitiation. Escape from this negative control is accomplished through the action of P-TEF (positive transcription elongation factor) which allows productive elongation. Fractionation studies have recently identified three components believed to be required to efficiently generate productive elongation complexes, P-TEFa, P-TEFb and factor 2 (Marshall and Price, 1995). P-TEFb was further purified and was shown to act after initiation (Marshall and Price, 1995), although the protein was not subject to detailed biochemical characterization.

The existence of two classes of transcription complexes differing in their elongation potential has also been demonstrated using the nucleoside analog, 5,6-dichloro-1- β -D-ribofuranosylbenzimidazole (DRB). The addition of DRB to mammalian cells in culture resulted in a 95% inhibition in the production of mature mRNA (Sehgal *et al.*, 1976). Nuclei isolated from cells pre-treated with DRB have increased production of short, capped transcripts while labeling of longer RNAs is decreased (Tamm and Kikuchi, 1979; Tamm *et al.*, 1980). Similarly, the short transcripts generated from viral templates in cells infected with SV40 (Laub *et al.*, 1980) and adenovirus (Fraser *et al.*, 1979) are enhanced, while longer transcripts are suppressed with DRB treatment. DRB also inhibits production of long transcripts but leaves shorter

products unaffected in injected *Xenopus* oocytes (Meulia *et al.*, 1993; Roberts and Bentley, 1992).

The carboxyl-terminal domain (CTD) of RNA polymerase II is phosphorylated during the transcription cycle at a time coincident with elongation regulation (Dahmus, 1994; Dahmus, 1995). The CTD can be phosphorylated by the kinase associated with the general transcription factor TFIH (Lu *et al.*, 1992; Serizawa *et al.*, 1992; Feaver *et al.*, 1991), and a CTD kinase activity is believed to be present in preinitiation complexes at several promoters (Peterson *et al.*, 1992; Kang and Dahmus, 1993). Research has been directed at identifying the CTD kinase, but despite the proposal of various candidate kinases, it appears that the relevant kinase has not yet been identified.

A kinase/cyclin pair (SRB10/11) is part of the holoenzyme form of yeast RNA polymerase II (Liao *et al.*, 1995). A number of other kinases, including casein kinase I and II (Zandomeni *et al.*, 1986; Cadena and Dahmus, 1987), DNA-dependent protein kinase (Dvir *et al.*, 1992), and a murine kinase related to *cdc2* and CDC28 (Cisek and Corden, 1989), are capable of phosphorylating the CTD. Also, the kinases CTD-K1 and CTD-K2 purified from HeLa cells (Payne and Dahmus, 1993), CTK1 from yeast (Lee and Greenleaf, 1989), and KI, KII, and KIII from *Aspergillus nidulans* (Stone and Reinberg, 1992) can all phosphorylate the CTD. It has been further suggested that the stress activated MAP kinases are involved in phosphorylating RNA polymerase II during heat shock (Venetianer *et al.*, 1995). While all of the above are serine/threonine kinases, there is one example of a tyrosine kinase, *c-abl*, that can phosphorylate the CTD (Baskaran *et al.*, 1993).

While phosphorylation of the CTD has been correlated with the elongation phase of transcription, none of the kinases described above have been shown to modify the functional properties of RNA polymerase II during elongation. Therefore, the identity of the kinase that operates in this control process remains unknown. The mechanism by which CTD phosphorylation induces the transition into productive elongation also remains to be determined, as does the role in elongation control of various proteins associated with RNA polymerase II in a

holoenzyme complex (Koleske and Young, 1995). Although several models for the involvement of CTD in elongation control have been proposed (Rasmussen and Lis, 1995), including tethering of the polymerase to the promoter by the unphosphorylated CTD, no conclusive evidence for one model is available. Further work is still needed to determine the fate of 5 polymerases that stop early, but do not enter productive elongation (Marshall and Price, 1992), and to define the interaction of RNA polymerases in early elongation complexes with termination factors, including "factor 2" (Xie and Price, 1996).

Thus, the role of termination factors and potential anti-termination factors, and the 10 regulation mechanisms that operate in the elongation control process remain to be clarified. Not only will the identification of these factors and their respective properties be of significant scientific interest, such discoveries would also have practical values beyond an understanding of transcriptional control mechanisms. For example, many viruses produce viral proteins that somehow interact with RNA polymerase II and facilitate the production of elongated viral transcripts, an essential step in the viral 'life cycle'. Therefore, the identification and 15 characterization of protein factors involved in productive elongation will likely yield benefits in the development of anti-viral strategies.

SUMMARY OF THE INVENTION

The present invention provides novel genes, proteins and related biological compositions developed for their ability to interact with RNA polymerase II and to control transcriptional elongation. The compositions of the present invention, which are based upon P-TEFb, are particularly beneficial as they also functionally interact with viral proteins, such as HIV Tat, that 20 have central roles in viral transcription elongation. Methods for identifying substances that alter transcription elongation are also provided by the invention, with particular emphasis on the identification of substances that inhibit the binding or functional interaction of viral proteins and P-TEFb, which substances are candidate anti-viral agents.

The invention was, in part, initially based upon the inventor's surprising discovery that the transcription elongation factor termed P-TEFb phosphorylates RNA polymerase II and controls the transition from abortive to productive transcription elongation. The inventor further discovered that human P-TEFb binds to the HIV transcriptional transactivating protein, Tat, and
5 that P-TEFb is the key host cell component that facilitates Tat-mediated viral mRNA elongation during HIV infection. From the initial findings, the inventor developed screening assays to identify substances that inhibit the interaction of human P-TEFb and viral proteins, such as HIV Tat, and will have utility as anti-viral agents.

10 Certain of the inventor's findings concern the cloning, for the first time, of each of the subunits that make up the *Drosophila* P-TEFb enzyme complex. The *Drosophila* cloning allowed the inventor to make the breakthrough in cloning the full length human counterpart P-TEFb subunit genes. The small, kinase subunit of human P-TEFb is herein identified as the product of a cDNA for which the nucleic acid sequence was known, but to which no known function had been ascribed. The large, cyclin-like subunit of human P-TEFb has been discovered by the present inventor and is disclosed for the first time in the present application. This invention therefore provides novel human biological components, including genes, proteins and purified holoenzymes, and also new screening methods based upon the use of the human P-TEFb enzyme complex.
15
20

P-TEFb is a key regulator of the process controlling the processivity of RNA polymerase II. The inventor has shown that P-TEFb can phosphorylate the CTD of pure RNA polymerase II. Furthermore, P-TEFb can phosphorylate the CTD of RNA polymerase II when the polymerase is in an early elongation complex. Both the function and kinase activity of P-
25 TEFb are blocked by the drugs DRB and H-8. P-TEFb is distinct from TFIIH because the two factors have no subunits in common, P-TEFb is more sensitive to DRB than TFIIH and, most importantly, TFIIH can not substitute functionally for P-TEFb. This invention discloses that phosphorylation of the CTD by P-TEFb controls the transition from abortive into productive elongation mode.
30

Both the human and the *Drosophila* small, or kinase, subunits are members of the CDC2-family of kinases and are cyclin-dependent kinases. The human small subunit associates with the activation domain of HIV-1 Tat, indicating that human P-TEFb is the Tat-associated kinase (TAK), previously poorly characterized and not purified prior to the present invention. An *in vitro* transcription assay demonstrates that the effect of Tat on transcription elongation requires P-TEFb and suggests the enhancement of transcriptional processivity by Tat is due to enhanced function of P-TEFb.

The present invention provides DNA segments, vectors and the like comprising at least one isolated gene, DNA segment or coding region that encodes a *Drosophila* P-TEFb kinase or large subunit protein, polypeptide, domain, peptide or any fusion protein thereof. Further provided are at least a first isolated gene, DNA segment or coding region that encodes a human P-TEFb large subunit protein, peptide, domain or derivative.

These aspects of the present invention may be described as follows: a DNA segment comprising an isolated coding region that encodes a substantially full length P-TEFb subunit, wherein the coding region is characterized as:

- a) encoding a substantially full length P-TEFb kinase subunit having the amino acid sequence of SEQ ID NO:2; or
- b) encoding a substantially full length P-TEFb large subunit that includes a contiguous sequence of at least about 7 amino acids from SEQ ID NO:4, SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50; or as a substantially full length coding region that hybridizes to the nucleotide sequence of SEQ ID NO:3, SEQ ID NO:43 or SEQ ID NO:48 under stringent hybridization conditions.

The term "substantially full length" as used herein, means that the genes and coding regions of the invention encode a substantially full length P-TEFb kinase or large subunit protein or polypeptide such that the subunit produced on expression of the gene or coding region includes each of the polypeptide regions or domains necessary to impart functional activity to the

expressed product. As disclosed herein, even subunits from the same species may vary in length and yet still have functional activity. For example, variations of the human P-TEFb large subunits are provided by the present invention which have differing lengths and yet still have biological activity. In particular, human large subunits of 696, 729 and 726 amino acids in length are provided hereby. Genes and coding regions that encode a protein of between about 600 and about 750, or preferably of between about 650 and about 750 amino acids in length are thus generally considered to be substantially full length genes or coding regions, although smaller genes and coding regions are by no means excluded from the present invention so long as they encode a protein that has biological activity when expressed.

10

It will also be understood that more variability in length will likely exist between genes and coding regions encoding P-TEFb subunits from different species. By way of example only, the *Drosophila* P-TEFb large subunits included within the invention encode a protein of about 1097 amino acids in length that has biological activity. A comparison of this protein with the human large subunit proteins disclosed herein reveals that there may be some considerable variability in the full length sequences of active subunits from different species. So long as a gene encodes a protein subunit that has biological activity as disclosed herein, this will be a "substantially full length" gene as this term is presently used.

20

The P-TEFb kinase subunits provided by and for use in the present invention also exemplify the concept of substantially full length and active proteins that may nonetheless have certain differences in sequence and actual length. The *Drosophila* kinase subunit provided hereby is about 404 amino acids in length, whereas the human subunit for use in the various methods and combined compositions of the invention is about 372 amino acids long. Each of these proteins have biological activity, as disclosed herein.

25

Although the concept of substantially full length sequences will be readily understood by those of ordinary skill in the art, another means for assessing the substantially full length nature of a gene, coding region or expressed protein of the invention is to analyze the terminal or near-terminal sequences of the biological components and to confirm that they generally correlate

30

with sequences at the termini of the sequences provided herein, or sequences proximal to such termini. By conducting such a comparison, one may identify substantially full length genes or coding regions that express biologically active proteins with even more variation in length.

5 Such a terminal sequence comparison is contemplated to be an effective means for identifying substantially full length biological components that may have inserted into their sequence additional coding or non-coding sequences (in the context of DNA) or additional polypeptide sequences. Working examples of this phenomenon are also provided herein, as a comparison of the human P-TEFb large subunit of SEQ ID NO:45 and SEQ ID NO:47 reveals
10 that the second encoded subunit contains additional polypeptide sequence that results from translation of an apparent intron in the DNA sequence. Genes encoding such introns and polypeptides including additional amino acid sequences are clearly encompassed within the present invention, and comparison of the substantially terminal sequences is an effective means for confirming the identity of longer sequences that are nonetheless P-TEFb coding regions or protein subunits.
15

In providing a DNA segment or vector that comprises an isolated gene or coding sequence that encodes a P-TEFb kinase subunit, protein or peptide, particularly a *Drosophila* P-TEFb kinase subunit, protein or peptide, the subunit may be generally characterized as:

- 20
- a) having an observed molecular weight of between about 42 kD and about 43 kD, generally as measured by gel filtration chromatography and SDS-PAGE (sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis PAGE); and having an actual molecular weight of about 47 kD, as calculated from the known protein sequence; and optionally characterized as
25
 - b) capable of forming a cyclin kinase pair with the P-TEFb large subunit protein, e.g., capable of binding to the *Drosophila* P-TEFb large subunit protein under suitable binding conditions, such as conditions using non-denaturing buffers of non-dissociating ionic strength and non-dissociating pH, which buffers generally
30

approximate in effect to the non-denaturing, non-dissociating cellular or nuclear conditions in which the P-TEFb protein complex exists in the natural state; and optionally characterized as

- c) being a cyclin-dependent kinase (CDK) that is capable of phosphorylating RNA polymerase II, and capable of phosphorylating RNA polymerase II when operatively combined with a P-TEFb large subunit protein, e.g., the kinase subunit is capable of binding to *Drosophila* RNA polymerase II in a manner and for a period of time effective to catalyze the transfer a phosphate group to RNA polymerase II from an available phosphate group donor molecule, such as adenosine triphosphate (ATP); wherein the RNA polymerase II is preferably phosphorylated on the carboxyl terminal domain (CTD) of the *Drosophila* large subunit of RNA polymerase II;

(i) more preferably, being identified as a DRB-sensitive cyclin-dependent kinase, wherein the phosphorylation of RNA polymerase II by the isolated P-TEFb kinase subunit protein, or the P-TEFb kinase subunit protein within the P-TEFb enzyme complex, is inhibited by an effective amount of DRB, and most preferably, wherein the phosphorylation of RNA polymerase II by the P-TEFb kinase subunit protein is inhibited by an effective amount of DRB and by an effective amount of H-8; and optionally characterized as

- d) being involved in the control of elongation by RNA polymerase II, and generally being capable of promoting transcription elongation when combined with a P-TEFb large subunit protein, e.g., capable of promoting proper elongation of mRNA transcripts when operatively combined with a P-TEFb large subunit to form a P-TEFb enzyme complex and wherein the P-TEFb enzyme complex is contacted with a functional *Drosophila* RNA polymerase II molecule in the presence of a DNA template and under conditions otherwise appropriate to result

in transcription elongation, *i.e.*, in the presence of effective amounts of nucleotides, ATP, other necessary co-factors and the like;

- (i) more preferably, wherein the capacity to promote transcription elongation when operatively combined with a P-TEFb large subunit protein, a functional RNA polymerase II and the necessary transcriptional elongation components, co-factors and precursors, is inhibited by an effective amount of DRB.

10 The genes and DNA segments preferably encode a substantially full length P-TEFb kinase subunit, protein or polypeptide that includes a contiguous amino acid sequence of at least about 24 amino acids, and more preferably, of at least about, 25, 27, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 200 amino acids or so from SEQ ID NO:2, or a biologically functional equivalent thereof. More preferably, the genes and DNA segments will encode a substantially full length P-TEFb kinase subunit having the amino acid sequence of SEQ ID NO:2, or a biologically functional equivalent thereof.
15

Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence of at least about 722 nucleotides, and more preferably, of at least about 725, 750, 800, 825, 850, 900 or so nucleotides from between position 115 and position 1326 or 1327 of SEQ ID NO:1, or a biologically functional equivalent thereof. More preferably, the isolated genes and DNA segments will comprise an isolated coding region having the nucleic acid sequence of the foregoing coding region SEQ ID NO:1, or a biologically functional equivalent thereof.

25 The human small kinase subunit, SEQ ID NO:6, comprises a single protein called "PITALRE", so called because of the presence of those amino acids in a characteristic location in the kinase subunit, which is encoded by the sequence presented in SEQ ID NO:5. It is a cyclin-dependent kinase of the cell division cycle 2 (CDC2) family of kinases.

In providing new combined compositions and new uses for the DNA segments herein discovered to encode the human P-TEFb kinase subunit protein (particularly combined compositions and uses in connection with DNA segments encoding a P-TEFb large subunit), the human P-TEFb kinase subunit gene is generally characterized as follows:

5

- a) encoding a human P-TEFb kinase subunit, protein or peptide that includes a contiguous amino acid sequence of at least about 6, 8, 10, 15, 20 or 24 amino acids or so from SEQ ID NO:6, or a biologically functional equivalent thereof; and optionally characterized as
- b) including a contiguous nucleic acid sequence of at least about 20-21 nucleotides or so, and more preferably, of at least about 30, 40, 50, 60, 72 or so nucleotides from the coding region of SEQ ID NO:5, or a biologically functional equivalent thereof.

10

15

Preferably, these genes and DNA segments will encode a human P-TEFb kinase subunit comprising, consisting essentially of, or having the contiguous amino acid sequence of SEQ ID NO:6, or a biologically functional equivalent thereof; or the genes and DNA segments will hybridize thereto under stringent hybridization conditions. These isolated genes and coding regions will therefore preferably include a contiguous nucleic acid sequence corresponding to substantially the full length coding region of SEQ ID NO:5, or a biologically functional equivalent thereof; or the genes and DNA segments will hybridize thereto under stringent hybridization conditions. Isolated genes and DNA segments having the nucleic acid sequence of SEQ ID NO:5 are just one example.

20

25
In providing new uses for the newly discovered human P-TEFb kinase subunit protein, particularly in combination with other P-TEFb subunits, proteins or peptides such that, for example, a recombinant human P-TEFb holoenzyme, as in Example 5, is provided, the human P-TEFb kinase subunit protein is also generally characterized as follows:

30

5 a) capable of forming a cyclin kinase pair with a human P-TEFb large subunit protein, or other large subunit protein, such that it is capable of binding to a human P-TEFb large subunit protein under suitable binding conditions, such as conditions using non-denaturing buffers of non-dissociating ionic strength and non-dissociating pH, which buffers generally approximate in effect to the non-denaturing, non-dissociating cellular or nuclear conditions in which the P-TEFb protein complex exists in the natural state;

10 b) being a cyclin-dependent kinase (CDK) that is capable of phosphorylating human RNA polymerase II, and capable of phosphorylating human RNA polymerase II when operatively combined with a human P-TEFb large subunit protein, *i.e.*, the kinase subunit is capable of binding to RNA polymerase II in a manner and for a period of time effective to catalyze the transfer a phosphate group to RNA polymerase II from an available phosphate group donor molecule, such as ATP; wherein the human RNA polymerase II is preferably phosphorylated on the carboxyl terminal domain (CTD) of the large subunit of RNA polymerase II;

15 (i) more preferably, being identified as a DRB-sensitive cyclin-dependent kinase, wherein the phosphorylation of human RNA polymerase II by the isolated human P-TEFb kinase subunit protein, or the P-TEFb kinase subunit protein within the human P-TEFb enzyme complex, is inhibited by an effective amount of DRB, and most preferably, wherein the phosphorylation of RNA polymerase II by the P-TEFb kinase subunit protein is inhibited by an effective amount of DRB and by an effective amount of H-8; and

20

25 c) being involved in the control of elongation by human RNA polymerase II, and generally being capable of promoting transcription elongation when combined with a human P-TEFb large subunit protein, *i.e.*, capable of promoting proper elongation of mRNA transcripts when operatively combined with a human

30

5

P-TEFb large subunit to form a P-TEFb enzyme complex and wherein the human P-TEFb enzyme complex is contacted with a functional human RNA polymerase II molecule in the presence of a DNA template and under conditions otherwise appropriate to result in transcription elongation, *i.e.*, in the presence of effective amounts of nucleotides, ATP, other necessary co-factors and the like;

10

(i) more preferably, wherein the capacity to promote transcription elongation when operatively combined with a human P-TEFb large subunit protein, a functional human RNA polymerase II and the necessary transcriptional elongation components, co-factors and precursors, is inhibited by an effective amount of DRB.

In certain embodiments, the DNA segments and coding regions may encode *Drosophila* and human P-TEFb kinase subunit peptides, for example from about 25 to about 30 or about 50 amino acids in length or so. Preferably, the DNA segments and coding sequences will encode a *Drosophila* P-TEFb kinase subunit protein of about 404 amino acids in length; or a human P-TEFb kinase subunit protein of about 372 amino acids in length, preferably where the human subunit is used in combination with other DNA segments, such that a recombinant human P-TEFb total enzyme is encoded.

20

The DNA segments and vectors of the present invention may comprise an isolated gene or coding region that encodes a *Drosophila* P-TEFb large subunit, protein or peptide. The *Drosophila* P-TEFb large subunit, protein or peptide may be generally characterized as:

25

- a) having a molecular weight of about 121 kD, generally as measured by gel filtration chromatography and also by calculating the molecular weight from the known protein sequence; and optionally characterized as
- b) capable of forming a cyclin kinase pair with the *Drosophila* P-TEFb kinase subunit protein, *i.e.*, being capable of binding to the *Drosophila* P-TEFb kinase

30

5

subunit protein under suitable binding conditions, such as conditions using non-denaturing buffers of non-dissociating ionic strength and non-dissociating pH, which buffers generally approximate to the non-denaturing, non-dissociating cellular or nuclear conditions in which the P-TEFb protein complex exists in the natural state; and optionally characterized as

10

- c) having cyclin protein-like sequence characteristics, such as the conserved cyclin box domain; and optionally characterized as
- d) capable of promoting transcription elongation when combined with a *Drosophila* P-TEFb kinase subunit protein, *i.e.*, capable of promoting proper elongation of mRNA transcripts when operatively combined with a *Drosophila* P-TEFb kinase subunit to form a P-TEFb enzyme complex and wherein the *Drosophila* P-TEFb enzyme complex is contacted with a functional *Drosophila* RNA polymerase II molecule in the presence of a DNA template and under conditions otherwise appropriate to result in transcription elongation, *i.e.*, in the presence of effective amounts of nucleotides, ATP, other necessary co-factors and the like,
 - (i) more preferably, wherein the capacity to promote transcription elongation when operatively combined with a *Drosophila* P-TEFb kinase subunit protein, a functional *Drosophila* RNA polymerase II and the necessary transcriptional elongation components, co-factors and precursors, is inhibited by an effective amount of DRB.

15
20

25

30

Certain genes and DNA segments preferably encode a substantially full length P-TEFb large subunit, protein or polypeptide that includes a contiguous amino acid sequence of at least about 7 amino acids, or more preferably, of at least about 8, 10, 12, 14, 16, 18, 20, 22 or 25 amino acids or so from SEQ ID NO:4, or a biologically functional equivalent thereof; or the genes and DNA segments will hybridize to such a coding sequence under stringent hybridization conditions. More preferably, the genes encode a P-TEFb large subunit, protein or peptide that

includes a contiguous amino acid sequence of at least about 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 200 amino acids or so from SEQ ID NO:4, or a biologically functional equivalent thereof; or the genes and DNA segments will hybridize to such a coding sequence under stringent hybridization conditions. Most preferably, these genes and DNA segments will encode
5 a P-TEFb large subunit having the amino acid sequence of SEQ ID NO:4, or a biologically functional equivalent thereof; or the genes and DNA segments will hybridize to such a coding sequence under stringent hybridization conditions.

These isolated genes and coding regions may include a contiguous nucleic acid sequence
10 of at least about 20-21 nucleotides or so, and more preferably, of at least about 30, 40, 50, 60 or 72 or so nucleotides from between position 716 and position 4054 of SEQ ID NO:3, or a biologically functionally equivalent thereof; or the genes and DNA segments will hybridize to such a coding sequence under stringent hybridization conditions. Preferably, these isolated genes and DNA segments will comprise an isolated coding region having the nucleic acid sequence of
15 SEQ ID NO:3, or a biologically functional equivalent thereof; or the genes and DNA segments will hybridize to such a coding sequence under stringent hybridization conditions.

Exemplary genes and DNA segments may also be characterized as encoding a substantially full length P-TEFb large subunit including a contiguous amino acid sequence of at least about 7 amino acids, or more preferably, of at least about 8, 10, 12, 14, 16, 18, 20, 22 or 25 amino acids or so from SEQ ID NO:4, or a biologically functional equivalent thereof; and as hybridizing to the nucleic acid sequence of SEQ ID NO:3 under stringent hybridization conditions.
20

25 In certain embodiments, the isolated DNA segments and coding regions may encode a P-TEFb large subunit peptide of from about 15 to about 30 or about 50 amino acids in length or so. Preferably, the DNA segments and coding regions encode a *Drosophila* P-TEFb large subunit protein of about 1113 amino acids in length.

The present invention provides several human P-TEFb large subunit genes, proteins and compositions. Methods of using the various compositions, for example, in the diagnosis and treatment of a viral infection, such as HIV, or cancer are also provided. Human P-TEFb large subunit proteins and peptides are generally characterized as:

5

- a) capable of forming a cyclin kinase pair with the human P-TEFb kinase subunit protein, *i.e.*, being capable of binding to the human P-TEFb kinase subunit protein of SEQ ID NO:6 under suitable binding conditions, such as conditions using non-denaturing buffers of non-dissociating ionic strength and non-dissociating pH, which buffers generally approximate to the non-denaturing, non-dissociating cellular or nuclear conditions in which the P-TEFb protein complex exists in the natural state; and
 - b) having cyclin protein-like sequence characteristics, such as the conserved cyclin box domain; and
 - c) being capable of promoting transcription elongation when combined with a human P-TEFb kinase subunit protein, *i.e.*, capable of promoting proper elongation of mRNA transcripts when operatively combined with a human P-TEFb kinase subunit to form a human P-TEFb enzyme complex and wherein the P-TEFb enzyme complex is contacted with a functional human RNA polymerase II molecule in the presence of a DNA template and under conditions otherwise appropriate to result in transcription elongation, *i.e.*, in the presence of effective amounts of nucleotides, ATP, other necessary co-factors and the like,
- 25
- (i) more preferably, wherein the capacity to promote transcription elongation when operatively combined with a human P-TEFb kinase subunit protein, a functional human RNA polymerase II and the necessary transcriptional elongation components, co-factors and precursors, is inhibited by an effective amount of DRB.
- 30

The human genes and DNA segments preferably encode a substantially full length human P-TEFb large subunit, protein or polypeptide that includes a contiguous amino acid sequence of at least about 6 or 7 amino acids or so, or more preferably, of at least about 8, 10, 12, 14, 16, 18, 5 20, 22 or 25 amino acids or so from SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50, or a biologically functional equivalent thereof; or the genes and DNA segments will hybridize to such a coding sequence under stringent hybridization conditions. More preferably, the genes encode a P-TEFb large subunit, protein or peptide that includes a contiguous amino acid sequence of at least about 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 200 amino acids or so from SEQ ID 10 NO:45, SEQ ID NO:47 or SEQ ID NO:50, or a biologically functional equivalent thereof; or the genes and DNA segments will hybridize to such a coding sequence under stringent hybridization conditions.

In certain preferred aspects, the human genes and DNA segments of the present invention will encode a P-TEFb large subunit having the amino acid sequence of SEQ ID NO:45, SEQ ID 15 NO:47 or SEQ ID NO:50, or a biologically functional equivalent thereof; or the genes and DNA segments will hybridize to such a nucleic acid segment under stringent hybridization conditions.

The isolated human genes and coding regions may include a contiguous nucleic acid sequence of at least about 20-21 nucleotides or so, and more preferably, of at least about 30, 40, 50, 60 or 72 or so nucleotides from the coding region of SEQ ID NO:43 or SEQ ID NO:48, or a biologically functionally equivalent thereof; or the genes and DNA segments will hybridize to such a coding sequence under stringent hybridization conditions. Preferably, the isolated genes 20 and DNA segments will comprise an isolated coding region having the nucleic acid sequence of SEQ ID NO:44, SEQ ID NO:46 or SEQ ID NO:49, or a biologically functional equivalent thereof; or the genes and DNA segments will hybridize thereto under stringent hybridization 25 conditions.

Exemplary human genes and DNA segments may also be characterized as encoding a substantially full length P-TEFb large subunit including a contiguous amino acid sequence of at 30

least about 7 amino acids, or more preferably, of at least about 8, 10, 12, 14, 16, 18, 20, 22 or 25 amino acids or so from SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50, or a biologically functional equivalent thereof; and as hybridizing to the nucleic acid sequence of SEQ ID NO:43 or SEQ ID NO:48 under stringent hybridization conditions.

5

The present invention also provides DNA segments comprising a first isolated gene or coding region that encodes a *Drosophila* or human P-TEFb kinase subunit, protein or polypeptide and a second isolated gene or coding region that encodes a corresponding *Drosophila* or human P-TEFb large subunit, protein or peptide.

10

As such, the present invention provides an expression system comprising:

- a) a first expression unit comprising, under the transcriptional control of a promoter, a first coding region that encodes a substantially full length P-TEFb kinase subunit that includes a contiguous sequence of at least about 7 amino acids from SEQ ID NO:2 or SEQ ID NO:6 and
- b) a second expression unit comprising, under the transcriptional control of a promoter, a second coding region that encodes a substantially full length P-TEFb large subunit that includes a contiguous sequence of at least about 7 amino acids from SEQ ID NO:4, SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50.

15
20

Such expression systems may comprise a first and second expression unit on a single expression vector; or a first and second expression unit on two distinct expression vectors. The expression systems may be advantageously comprised within a recombinant host cell.

25 Such host cells will generally express a substantially full length P-TEFb kinase subunit and a substantially full length P-TEFb large subunit. These cells will therefore produce P-TEFb enzyme complexes comprising subunits generally of the same species. However, cross-species P-TEFb enzyme complexes, e.g., those that comprise one *Drosophila* subunit and one human subunit, are also contemplated.

30

Isolated, functional P-TEFb enzyme complexes thus form another aspect of the present invention. The *Drosophila* P-TEFb enzyme complex is generally characterized as:

- a) comprising at least two subunits in operative association, a first, kinase subunit having a molecular weight of about 42 kD, and a second, large subunit (or cyclin-related subunit) having a molecular weight of about 121 kD;
- b) being naturally localized to the cell nucleus and capable of interacting with and phosphorylating *Drosophila* RNA polymerase II present in an early elongation complex under suitable cellular conditions, or under suitable *in vitro* binding conditions that effectively duplicate the non-denaturing, non-dissociating cellular transcription elongation conditions;
- c) being capable of phosphorylating *Drosophila* RNA polymerase II; and preferably, capable of phosphorylating *Drosophila* RNA polymerase II on the carboxyl terminal domain (CTD) of the large subunit of RNA polymerase II; and more preferably, wherein the capacity to phosphorylate *Drosophila* RNA polymerase II is inhibited by an effective amount of DRB; and
- d) being capable of promoting transcriptional elongation, which is also termed promoting the transition from an abortive to a productive elongation mode, *i.e.*, being capable of promoting proper or productive elongation of mRNA transcripts (rather than abortive elongation) when operatively combined with a functional *Drosophila* RNA polymerase II molecule in the presence of a DNA template and under conditions otherwise appropriate to result in transcription elongation, *i.e.*, in the presence of effective amounts of nucleotides, ATP, other necessary co-factors and the like; and preferably, wherein the capacity to promote transcriptional elongation is inhibited by an effective amount of DRB.

The *human* P-TEFb enzyme complex is generally characterized as:

- PCT/EP/2015/060088
15
20
- 25
- a) comprising at least two subunits in operative association, a first, kinase subunit preferably comprising the sequence of SEQ ID NO:6, and a second, large or cyclin-related subunit preferably comprising the sequence of SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50;
 - b) being naturally localized to the human cell nucleus and capable of interacting with and phosphorylating human RNA polymerase II present in an early elongation complex under suitable cellular conditions, or under suitable *in vitro* binding conditions that effectively duplicate the non-denaturing, non-dissociating cellular transcription elongation conditions;
 - c) being capable of phosphorylating human RNA polymerase II; and preferably, capable of phosphorylating human RNA polymerase II on the carboxyl terminal domain (CTD) of the large subunit of RNA polymerase II; and more preferably, wherein the capacity to phosphorylate human RNA polymerase II is inhibited by an effective amount of DRB; and
 - d) being capable of promoting transcriptional elongation, which is also termed promoting the transition from an abortive to a productive elongation mode, *i.e.*, being capable of promoting proper or productive elongation of mRNA transcripts (rather than abortive elongation) when operatively combined with a functional human RNA polymerase II molecule in the presence of a DNA template and under conditions otherwise appropriate to result in transcription elongation, *i.e.*, in the presence of effective amounts of nucleotides, ATP, other necessary co-factors and the like; and preferably, wherein the capacity to promote transcriptional elongation is inhibited by an effective amount of DRB.

DNA segments and isolated genes may also be manipulated to encode a P-TEFb subunit fusion protein or polypeptide construct in which at least one P-TEFb subunit, protein,

polypeptide or even peptide is operatively attached to a second coding region that encodes a selected peptide or protein sequence. The combination of P-TEFb subunit sequences, including human subunits, with selected antigenic amino acid sequences; selected non-antigenic carrier amino acid sequences, for use in immunization; selected adjuvant sequences; amino acid sequences with specific binding affinity for a selected molecule; amino acid sequences that form an active DNA binding or transactivation domain are particularly contemplated. Certain fusion proteins may be linked together *via* a protease-sensitive peptide linker, allowing subsequent easy separation.

The DNA segments intended for use in expression will be operatively positioned under the control of, *i. e.*, downstream from, a promoter that directs expression of a P-TEFb subunit, protein or polypeptide in a desired host cell, such as *E. coli*, or in certain other preferred embodiments in an insect, mammalian or human cell. The promoter may be a recombinant promoter, or a promoter naturally associated with P-TEFb. Recombinant vectors, including baculoviral vectors, thus form another aspect of the present invention. The recombinant vectors may express a *Drosophila* P-TEFb kinase subunit, protein or polypeptide and a *Drosophila* P-TEFb large subunit, protein or polypeptide. Recombinant vectors expressing human P-TEFb large subunit proteins or polypeptides, human P-TEFb kinase subunits in combination with large subunits or even human kinase subunits for various uses are also provided.

Although sequences encoding substantially full length subunits are preferred, the invention further provides nucleic acid probes and primers and other nucleic acid segments, including those characterized as including:

- a) comprising a sequence region that consists of at least about 120, 150, 200 or so contiguous nucleotides that have the same sequence as, or are complementary to, about 120, 150, 200 or so contiguous nucleotides selected from any region of SEQ ID NO:3; or

5

- b) a nucleic acid segment of from about 120, 150, 200 or so to about 20,000 nucleotides in length that hybridizes to any region of the nucleic acid segment of SEQ ID NO:3, or the complement thereof under standard hybridization conditions, and particularly under hybridization conditions in which the *Drosophila* sequence will not bind to the human kinase subunit sequence.

10

However, in defined regions of the sequence, e.g., those with less conservation, segments of each of SEQ ID NO:3, or the complement thereof, may variously be about 20, 25, 30, 50, 100, 200, 500, or 1000 or so nucleotides in length, up to and including full length sequences, or even longer, as may be achieved by duplication of certain regions. Where the sequence of SEQ ID NO:3 is concerned, sequences of at least about 2000, 3000, 4000 nucleotides of SEQ ID NO:3, or complements thereof are provided, up to and including the full length sequence of 4328 contiguous nucleotides of SEQ ID NO:3, or the complement thereof.

15
20
25

In addition, the invention provides nucleic acid probes and primers and other nucleic acid segments, including those characterized as including:

- a) comprising a sequence region that consists of at least about 360, 400, 450 or so contiguous nucleotides that have the same sequence as, or are complementary to, about 360, 400, 450 or so contiguous nucleotides selected from any region of SEQ ID NO:43, SEQ ID NO:46 or SEQ ID NO:48; or
- b) a nucleic acid segment of from about 360, 400, 450 or so to about 20,000 nucleotides in length that hybridizes to any region of the nucleic acid segment of SEQ ID NO:43, SEQ ID NO:46 or SEQ ID NO:48, or the complement thereof under standard hybridization conditions.

25

Again, in certain defined regions of the sequences, e.g. those with less conservation as defined herein, segments of each of SEQ ID NO:43, SEQ ID NO:46 or SEQ ID NO:48, or the complements thereof, may variously be about 20, 25, 30, 50, 100, 200, 500, or 1000 or so

nucleotides in length, up to and including full length sequences, or even longer, as may be achieved by duplication of certain regions.

It will be readily understood that what is meant by "defined regions of the sequence, e.g.,
5 those with less conservation" are contiguous stretches of at least about 20 nucleotides that are not identical or complimentary to known nucleic acid sequences. Such defined regions include, for example, positions 1-258, 320-345 and 1244-1457 of SEQ ID NO:1; positions 587-964, 1156-1711, 1764-3287, 3460-3775 and 3800-4328 of SEQ ID NO:3; and preferably, positions 1-244, 297-546, 867-1142, 1895-2331, 2821-2890, 3341-3442, 3953-3860 and 4491-4528 of SEQ ID
10 NO:43; and positions 1-209, 418-667, 919-1031, 2045-2164 and 2219-2360 of SEQ ID NO:48.

Where the sequence of SEQ ID NO:43, SEQ ID NO:46 or SEQ ID NO:48 is concerned, sequences of at least about 2000, 3000, 4000 nucleotides of SEQ ID NO:3, or complements thereof are provided, up to and including the full length sequence of 4528 contiguous nucleotides of SEQ ID NO:43, including the full length sequence of 2190 contiguous nucleotides of SEQ ID NO:46 or including the full length sequence of 2360 contiguous nucleotides of SEQ ID NO:48, or the complement thereof.
15

Any segment may be combined into a DNA segment or vector of up to about 30,000, about 20,000, or about 15,000 base pairs in length. Segments of up to about 20,000, 15,000, and 10,000 basepairs in length will generally be preferred, and segments of up to about 5,000 basepairs in length are also provided.
20

The nucleic acids of the present invention may also be DNA segments or RNA segments.
25 Nucleic acid detection kits are also provided.

The present invention further provides recombinant host cells comprising at least one DNA segment or vector that comprises an isolated gene or coding region that encodes a *Drosophila* or human P-TEFb subunit protein, polypeptide, domain or any fusion protein thereof.

Prokaryotic host cells, such as *E. coli*, are provided, as are eukaryotic host cells, such as insect cells and mammalian cells.

The recombinant host cells may further comprises an operative HIV Tat protein or active fragment thereof. Such recombinant host cells may be provided with the HIV Tat protein or peptide *in vitro*, for example, to test P-TEFb subunit protein and HIV Tat interactions, or may naturally express HIV Tat, including cells provided with P-TEFb subunit proteins, peptides or domains *in vivo* and *in vitro*.

The recombinant host cells of the present invention preferably have one or more DNA segments introduced into them by means of a recombinant vector, and preferably express the DNA segment to produce the encoded P-TEFb subunit, protein or polypeptide. The recombinant host cells may express a P-TEFb subunit, protein or polypeptide including a contiguous amino acid sequence of at least about 7 amino acids from SEQ ID NO:2 or SEQ ID NO:4, and preferably express a P-TEFb subunit, protein or polypeptide having the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:4.

More preferably, the recombinant host cells will express human P-TEFb large subunits, proteins or peptides that include a contiguous amino acid sequence of at least about 7 amino acids from SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50, and preferably express P-TEFb subunits having the amino acid sequence of SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50. Cells expressing any of the foregoing large subunits in combination with human P-TEFb kinase subunits, proteins or peptides, preferably those that include a contiguous amino acid sequence from SEQ ID NO:6, are also provided.

The recombinant host cells of the present invention may express a P-TEFb subunit, protein or polypeptide fusion protein in which a contiguous P-TEFb subunit amino acid sequence is operatively attached to a selected peptide or protein sequence. Also provided are recombinant host cells that express a *Drosophila* or human P-TEFb kinase subunit and a *Drosophila* or human P-TEFb large subunit. Cells that allow the production of human holoenzymes are preferred.

Methods for detecting P-TEFb nucleic acids in cells or samples are also provided, and generally comprise obtaining sample nucleic acids from a sample suspected of containing P-TEFb nucleic acids, contacting the sample nucleic acids with a nucleic acid segment that

5 encodes a P-TEFb subunit, protein or polypeptide, preferably a human P-TEFb subunit, protein or polypeptide, under conditions effective to allow hybridization of substantially complementary nucleic acids, and detecting the hybridized complementary nucleic acids thus formed. Use of nucleic acid segments that comprise a discriminating contiguous sequence from the sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:43, SEQ ID NO:48, SEQ ID NO:44,

10 SEQ ID NO:46 or SEQ ID NO:49 are preferred, with discriminating nucleic acid sequences from the sequence of SEQ ID NO:43, SEQ ID NO:48, SEQ ID NO:44, SEQ ID NO:46 or SEQ ID NO:49 being more preferred.

As the P-TEFb enzyme complex is herein shown to be essential to the transcription elongation process in eukaryotic, mammalian and human cells, it is envisioned that the levels of P-TEFb may correlate with various diseases, such as cancer. Accordingly, the present invention further provides methods of determining the levels of P-TEFb in cells or tissue samples, including tumor cells and samples, which generally comprise obtaining a biological sample suspected of containing P-TEFb; contacting the sample with a biological reagent that detects P-TEFb, under conditions effective to allow detection; and determining the level of P-TEFb detected. The "biological reagent that detects P-TEFb" may be a nucleic acid segment that encodes a human P-TEFb subunit, protein or polypeptide or it may be an antibody that has specific binding affinity for a human P-TEFb subunit, protein or polypeptide.

It is also contemplated that the "type" of P-TEFb, *i.e.*, the presence of one or more P-TEFb proteins or mutants thereof, may correlate with diseases, such as cancer. Accordingly, the invention even further provides methods of determining whether P-TEFb mutants or one or more different P-TEFb proteins are present in cells or tissue samples. The methods generally comprise obtaining a biological sample suspected of containing a mutant P-TEFb or different P-TEFb protein; contacting the sample with a biological reagent capable of detecting a mutant,

distinct from a wild type, P-TEFb or a different, or second, P-TEFb protein, under conditions effective to allow differential detection; and determining whether another P-TEFb protein or a P-TEFb mutant is present.

5 The "biological reagent that detects mutant P-TEFb" will generally be a nucleic acid segment or gene, or an antibody, that has specificity for the mutant sequence or protein in preference to the wild type sequence or protein, allowing effective differentiation between the two, as may be used in diagnostic tests for cancer cells or patients.

10 The "biological reagent that detects different P-TEFb" will generally be a nucleic acid segment or gene, or an antibody, that has specificity for another wild type sequence or protein in preference to the first wild type sequence or protein, allowing effective differentiation between the two, as may be used in diagnostic tests for cancer cells or patients. The use of nucleic acid segments, probes or primers that differentiate between the three different human P-TEFb large subunit proteins provided herein, *i.e.*, proteins having the amino acid sequence of SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50, is particularly contemplated.

15 The invention therefore also includes the provision of DNA segments, vectors, genes and coding sequence regions that encode *Drosophila* or human P-TEFb proteins, polypeptides, domains, peptides or any fusion protein thereof, where the P-TEFb protein element comprises at least one mutation in comparison to the wild type sequence. The mutation may be deliberately introduced by the hand of man, for example, in order to test the function of the changed amino acid, *e.g.*, in Tat or RNA polymerase II binding, and/or other functions. The mutation may be also be discovered in the natural population, as may be connected with dysfunction or disease.

20 Once a correlation between the levels of P-TEFb and a disease, such as cancer, has been confirmed, the present invention further provides methods for diagnosing said disease in other patients. Diagnostically, the present invention then provides methods for identifying a patient having or at risk for developing, for example, cancer; the methods comprising determining the type or amount of P-TEFb present within a biological sample from the patient, wherein the

presence of a type or amount of P-TEFb different to the type or amount of P-TEFb present in a corresponding sample from a normal subject, is indicative of a patient having or at risk for developing the disease.

5 Methods of using DNA segments that include an isolated P-TEFb subunit gene or coding region, including human DNA segments, are provided, wherein the methods comprise expressing a P-TEFb subunit DNA segment in a recombinant host cell and collecting the P-TEFb subunit protein, polypeptide, domain or fusion protein expressed by the cell. This method may be represented by the steps of:

- 10
- a) preparing a recombinant vector in which a P-TEFb subunit-encoding DNA segment, preferably a human DNA segment, is positioned under the control of a promoter;
 - b) introducing the recombinant vector into a recombinant host cell;
 - c) culturing the recombinant host cell under conditions effective to allow expression of an encoded P-TEFb subunit protein, polypeptide, domain or fusion protein; and
 - d) collecting the expressed P-TEFb subunit protein, polypeptide, domain or fusion protein.
- 15
- 20

The invention further provides recombinant P-TEFb subunit polypeptides, proteins and fusion proteins, preferably of human origin, prepared at levels that could not previously be obtained prior to the present invention. The methods comprise expressing a gene encoding a P-TEFb subunit polypeptide, protein or fusion protein in a recombinant host cell and purifying the expressed polypeptide, protein or fusion protein away from total recombinant host cell components to prepare between about 100 µg and about 1000 mg of a recombinant P-TEFb subunit polypeptide, protein or fusion protein. With scale up, the inventor contemplates that 10-fold increases can be achieved yielding up to about 10 g of recombinant P-TEFb proteins.

The invention also provides compositions comprising isolated P-TEFb subunit peptides, proteins or fusion proteins in all amounts between about 100 µg and about 1000 mg, such as between about 500 µg and about 100 mg.

5 P-TEFb fusion proteins or constructs including P-TEFb subunit, protein or polypeptide sequences operatively attached to distinct, selected amino acid sequences, such as selected antigenic amino acid sequences, amino acid sequences with selected binding affinity, and DNA binding or transactivation amino acid sequences, are also encompassed within the invention. Particularly, P-TEFb subunit, protein or polypeptide sequences operatively attached to

10 glutathione-S-transferase amino acid sequences are provided. Fusion proteins with selectively-cleavable bonds are also provided.

15 Compositions comprising isolated and purified P-TEFb kinase subunit proteins, polypeptides or fusion proteins that include a contiguous amino acid sequence of at least about 24 amino acids, and more preferably, of at least about, 25, 27, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 200 amino acids or so from SEQ ID NO:2, or a biologically functional equivalent thereof, are also provided. P-TEFb kinase subunit proteins, polypeptides or fusion proteins having the amino acid sequence of SEQ ID NO:2 or a biologically functional equivalent thereof are preferred.

20 Also provided are isolated and purified P-TEFb large subunit proteins, polypeptides or fusion proteins that include a contiguous amino acid sequence of at least about 6 or 7 or so amino acids, and more preferably, of at least about, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150 or 200 amino acids or so from SEQ ID NO:4, or a biologically functional equivalent thereof. P-TEFb large subunit proteins, polypeptides or fusion proteins having the amino acid sequence of SEQ ID NO:4 or a biologically functional equivalent thereof are more preferred.

Still more preferred aspects of this invention are isolated and purified human P-TEFb large subunit proteins, polypeptides or fusion proteins. These components generally include a contiguous amino acid sequence of at least about 6 or 7 or so amino acids, and more preferably, of at least about, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150 or 5 200 amino acids or so from SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50, or a biologically functional equivalent thereof. P-TEFb large subunit proteins, polypeptides or fusion proteins having the amino acid sequence of SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50, or a biologically functional equivalent thereof, are even more preferred still.

10 Further provided are compositions comprising isolated and purified active *Drosophila* P-TEFb enzyme complexes comprising P-TEFb kinase subunits in operative association with P-TEFb large subunits, preferably at levels that could not be previously obtained. Further provided are compositions comprising isolated and purified active human P-TEFb enzyme complexes comprising P-TEFb kinase subunits in operative association with P-TEFb large 15 subunits, preferably at levels that could not be previously obtained.

The invention further provides compositions comprising an HIV Tat protein in combination with an active P-TEFb enzyme complex comprising P-TEFb kinase subunits in operative association with P-TEFb large subunits. The complex is preferably a human complex.

20 The present invention provides P-TEFb immunodetection reagents. The immunodetection reagents may be characterized as:

- a) an antibody that has immunospecificity for a *Drosophila* P-TEFb kinase subunit protein, preferably a protein of at least about 7 amino acids from SEQ ID NO:2, or more preferably, a protein of SEQ ID NO:2; an antibody that has immunospecificity for a *Drosophila* P-TEFb large subunit protein, preferably a protein of at least about 7 amino acids from SEQ ID NO:4, or more preferably, a protein of SEQ ID NO:4; or an antibody that has immunospecificity for a human P-TEFb large subunit protein, preferably a protein of at least about 7 amino acids

5

- from SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50, or more preferably, a protein of SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50; any of which antibodies may be operatively attached to a detectable label; or
- b) an antibody that has immunospecificity for a human P-TEFb kinase subunit protein, preferably a protein of SEQ ID NO:6, wherein the antibody is operatively attached to a detectable label.

10 The immunodetection reagent comprising an antibody that has immunospecificity for a human P-TEFb large subunit protein, may further be attached to a detectable label. The detectable labels for use in the present invention may be a radioactive label, a fluorescent label, biotin, avidin or an enzyme that will generate a detectable product upon contact with an appropriate substrate, which is preferably a colored product. The antibodies for use in the immunodetection reagents of the present invention are preferably monoclonal antibodies.

15
20
25

The present invention provides immunodetection kits, certain of which may be described as comprising:

- a) a Tat composition comprising a purified HIV Tat protein;
- b) a P-TEFb composition comprising a purified human P-TEFb subunit; and
- c) an immunodetection means.

25 The P-TEFb composition may comprise a purified human P-TEFb kinase subunit, a purified human P-TEFb large subunit, or a purified human P-TEFb enzyme complex comprising a P-TEFb kinase subunit bound to a P-TEFb large subunit.

30 The immunodetection kits of the present invention provide a variety of immunodetection means. The immunodetection means may be a detectable label that is operatively attached to the

HIV Tat protein. In other kits, the immunodetection means may be a first anti-Tat antibody that binds to the HIV Tat protein, preferably wherein the first anti-Tat antibody is operatively attached to a detectable label. Additionally, the immunodetection means may be a detectable label that is operatively attached to the human P-TEFb subunit.

5

Alternatively, the immunodetection means may be a first anti-P-TEFb antibody that binds to a human P-TEFb subunit. The invention further provides kits wherein the first anti-P-TEFb antibody binds to human P-TEFb kinase subunits, human P-TEFb large subunits, or to intact human P-TEFb enzyme complexes. Further, the first anti-P-TEFb antibody may be operatively attached to a detectable label.

10

Certain other kits may comprise a first anti-Tat antibody or a first anti-P-TEFb antibody, wherein the immunodetection means is a detectable label that is operatively attached to a second antibody that has binding affinity for the first antibody. In such kits, the HIV Tat protein or the human P-TEFb subunit may be bound to a solid support.

15

The present invention further advantageously provides methods for identifying genes that encode P-TEFb subunit proteins from a desired species based upon the "two hybrid screening system". The methods rely on the use of one defined subunit to clone a matching subunit of the desired species, such that one of the binding pair (kinase or large subunit) is known and the other is identified by practicing the method. Accordingly, where one desires to clone a P-TEFb large subunit protein one uses the "complementary" P-TEFb kinase subunit protein in the screen. That is the meaning of "complementary" as used in the following methodological description. The methods may thus be characterized as comprising the steps of:

20

- a) obtaining a first DNA segment comprising a candidate P-TEFb subunit gene from a desired species; the first DNA segment expressing a first fusion protein comprising a transcriptional transactivating domain operatively attached to the candidate P-TEFb subunit protein encoded by the candidate gene;

25

- PCT/US2012/050111
- 15
- 20
- 25
- b) obtaining a second DNA segment that expresses a second fusion protein comprising the complementary P-TEFb subunit protein operatively attached to a DNA binding domain that binds to a defined nucleic acid sequence;
- 5 c) providing the first and second DNA segments to a eukaryotic host cell that comprises a marker gene operatively positioned downstream of the defined nucleic acid sequence; and
- 10 d) identifying a eukaryotic host cell that expresses the marker gene, thereby identifying the candidate gene as a gene of the desired species that encodes a P-TEFb subunit protein.

The methods generally further comprise isolating the identified candidate P-TEFb subunit gene from the first DNA segment within the eukaryotic host cell. U.S. Patent 5,667,973 is specifically incorporated herein by reference for the purposes of providing further details concerning the execution of screening methods based upon this general technique.

The transcriptional transactivating domains used in the present invention may be the GAL4 or VP16 transcriptional transactivating domain. The fusion protein may comprise a GAL4 DNA binding domain, wherein the defined nucleic acid sequence comprises a GAL4 binding domain recognition sequence, or a lexA DNA binding domain, wherein the defined nucleic acid sequence comprises a lexO binding site sequence. In the methods, the eukaryotic host cell may be a yeast host cell (yeast two hybrid system) or a mammalian host cell.

In the two hybrid system methods of the present invention, marker genes preferred for use are chloramphenicol acetyltransferase, β -galactosidase, green fluorescent protein, β -glucuronidase or the luciferase gene, preferably the β -galactosidase gene.

A further explanation of the two hybrid system cloning method for identifying a gene of a desired species that encodes a P-TEFb subunit protein is that it generally operatively comprises the steps of:

- 5 a) obtaining a plurality of first DNA segments comprising a plurality of candidate genes of the desired species;
- b) obtaining multiple copies of the second DNA segment;
- 10 c) providing the plurality of first DNA segments and multiple copies of the second DNA segments to a population of eukaryotic host cells in an amount sufficient to provide about one first DNA segment and at least about one second DNA segment to each host cell in the population;
- d) culturing the population of cells under conditions and for a period of time effective to allow marker gene expression; and
- e) detecting a host cell from the population that expresses the marker gene, thereby identifying the presence in the cell of a first DNA segment that comprises a candidate gene of the desired species that encodes a P-TEFb subunit protein.

The method also generally further comprises isolating the detected cell of step (e) free from the population of cells, and isolating the candidate gene of the desired species from the first DNA segment within the cell.

25 Although human P-TEFb-encoding nucleic acids are directly provided by the present invention, the invention also enables the identification and use of additional P-TEFb-encoding nucleic acid segments of any defined species. Methods for obtaining such sequences generally comprise the steps of:

30

- 5
- a) obtaining at least one isolated nucleic acid segment designed to hybridize to a P-TEFb subunit gene of the desired species;
 - b) contacting a population of nucleic acids of the desired species with the isolated nucleic acid segment under conditions effective to allow hybridization of the isolated nucleic acid segment to P-TEFb subunit nucleic acids within the population of nucleic acids of the desired species; and
 - 10 c) identifying a nucleic acid segment of the desired species that hybridizes to the isolated nucleic acid segment, preferably free from the population of nucleic acids of the desired species.

15

The methods of identifying nucleic acids of defined species that encode P-TEFb subunit proteins may utilize a single isolated nucleic acid segment designed to hybridize to a P-TEFb subunit gene of the desired species, wherein the population of nucleic acids of the desired species are contacted with the single isolated nucleic acid segment under conditions in which the single isolated nucleic acid sequence will bind to the large subunit sequence from the desired species. The single isolated nucleic acid segment may be designed from knowledge of the *Drosophila* or human genes provided by the present invention, or may be a probe sequence designed from a peptide purified from the desired species.

20

25

The methods of identifying P-TEFb subunit-encoding nucleic acids of a desired species may also utilize a pair of isolated nucleic acid segments or probes designed to hybridize to spatially distant sequences from a P-TEFb subunit gene of the desired species, wherein a polymerase chain reaction is conducted to amplify the P-TEFb subunit gene located between the spatially distant sequences, using protocols that are known to those of skill in the art. The probes may again be designed from the *Drosophila* and human sequences of the present invention or may be designed from peptide sequences of the desired species.

The present invention also provides methods of identifying a nucleic acid segment of a desired species that comprises an isolated coding region that encodes a P-TEFb subunit, protein or peptide wherein the isolated nucleic acid segment is designed from an analysis of at least one peptide sequence obtained from a P-TEFb subunit protein of the desired species. The peptide sequence may be an N-terminal or internal peptide sequence. In certain aspects of the present invention, the peptide sequence may be obtained by hydrolyzing a purified P-TEFb subunit protein from the desired species to form peptides and sequencing a peptide there formed. The method of obtaining the peptide sequence may comprise the steps of:

- 10 a) purifying a P-TEFb subunit protein from the desired species, preferably by fractionating a cellular extract comprising a P-TEFb subunit protein of the desired species and obtaining a fraction enriched for the subunit;
- 15 b) hydrolyzing the purified P-TEFb subunit protein to form a population of peptides;
- 20 c) isolating a single peptide from the population of peptides; and
- 25 d) sequencing the isolated peptide.

The present invention further provides methods for preparing purified P-TEFb subunit proteins of a desired species comprising the steps of:

- 30 a) obtaining a nuclear extract from a cell of the desired species;
- 35 b) subjecting the extract to fractionation *via* column chromatography, using a series of chromatography columns, such as phosphocellulose (P-11), DEAE cellulose, phenyl sepharose, hydroxylapatite, MonoQ, MonoS and the like; and preferably, subjecting the extract to fractionation *via* column chromatography using, in sequence, a phosphocellulose, phenyl sepharose, hydroxylapatite, MonoQ and MonoS column; and

- c) obtaining a fraction comprising the P-TEFb subunit protein substantially free from other protein components.

5 The invention also provides methods of preparing purified P-TEFb subunit proteins of a desired species that comprise the steps of:

- a) obtaining a nuclear extract from the desired species;
- b) subjecting the extract to fractionation using an antibody that binds to a known P-TEFb subunit protein;
- c) purifying a fraction comprising the desired P-TEFb subunit protein in combination with the known P-TEFb subunit protein; and
- d) separating the desired P-TEFb subunit protein free from the known P-TEFb subunit protein.

15 The antibodies used for preparing purified P-TEFb subunit proteins may be immobilized on a solid support, with the extract fractionated by applying the extract to the solid support. More preferably, the antibodies are monoclonal antibodies. These antibodies may be prepared by immunizing an animal with a P-TEFb subunit, protein or peptide, and collecting the resultant antibodies.

20 The present invention thus provides nucleic acid segments of any desired species that comprise an isolated coding region or gene that encodes a P-TEFb subunit protein of the desired species. These nucleic acid segments may also be comprised within a recombinant vector, which may be comprised within a recombinant host cell.

Compositions comprising purified P-TEFb subunit proteins of any desired species are further provided. The compositions may be obtained from cells that naturally express the P-TEFb subunits, or obtained from a recombinant cell that has been engineered to express the P-TEFb subunit of the desired species. The compositions may be prepared by fractionating a cell extract comprising a P-TEFb subunit protein and purifying the subunit protein away from total cell components. The fractionation may comprise column chromatography, preferably affinity column chromatography, and more preferably immunoaffinity column chromatography using a column comprising an antibody that binds to the P-TEFb subunit protein of the desired species.

The compositions comprising purified P-TEFb subunit proteins of a desired species may further comprise the other P-TEFb subunit protein of the binding pair, and may still further comprise a viral protein, such as a species-specific protein equivalent of the HIV Tat protein.

In particularly useful embodiments, this invention also provides methods of assaying for P-TEFb. One of the methods comprises testing a composition suspected of containing P-TEFb for the ability to phosphorylate RNA polymerase II, wherein phosphorylation is indicative of a composition comprising at least a P-TEFb kinase subunit, and preferably is indicative of a composition comprising a functional P-TEFb holoenzyme. The method may be described as comprising the steps of:

- a) admixing (i) a test composition suspected of containing P-TEFb, (ii) a composition comprising at least the carboxyl terminal domain (CTD) of the large subunit of RNA polymerase II, and (iii) an effective phosphate donor compound comprising a labeled phosphate group; and
- b) determining the ability of the test composition to catalyze the transfer of the labeled phosphate group to the carboxyl terminal domain (CTD) of RNA polymerase II.

The effective phosphate donor compound is preferably ATP that comprises a ^{32}P - or ^{33}P -labeled terminal phosphate group, although GTP also functions in this regard.

5 The invention also provides a phosphorylation-based method for identifying a candidate transcriptional inhibitor, comprising preparing a P-TEFb composition comprising at least a P-TEFb kinase subunit, and preferably comprising a functional P-TEFb holoenzyme, and testing the candidate inhibitor for the ability to inhibit P-TEFb-mediated phosphorylation of RNA polymerase II, wherein inhibition of phosphorylation is indicative of a candidate transcriptional inhibitor. The method may be described as comprising the steps of:

10

- a) obtaining a P-TEFb composition comprising at least a P-TEFb kinase subunit, and preferably comprising a functional P-TEFb holoenzyme;
- b) obtaining an RNA polymerase II composition comprising at least the carboxyl terminal domain (CTD) of the large subunit of RNA polymerase II;
- c) admixing the P-TEFb composition with the RNA polymerase II composition and an effective phosphate donor compound comprising a labeled phosphate group; and
- d) determining the ability of the P-TEFb composition to transfer the labeled phosphate group to the RNA polymerase II composition in the presence of the candidate transcriptional inhibitor and in the absence of the candidate transcriptional inhibitor, wherein a reduction in the amount of labeled phosphate transferred to RNA polymerase II in the presence of the candidate is indicative of a positive candidate transcriptional inhibitor.

25

15

20

25

30

The P-TEFb composition may comprise a *Drosophila* P-TEFb composition or a human P-TEFb composition. The P-TEFb composition will preferably comprise a P-TEFb enzyme complex that has transcription elongation promoting activity. The terms "a P-TEFb enzyme

complex that has transcription elongation promoting activity", or "a transcriptionally active P-TEFb enzyme complex" is a complex, generally comprising at least a kinase subunit and a large, cyclin-like subunit, which facilitates transcriptional elongation by RNA polymerase II, or removes a previous 'block' that was preventing proper transcriptional elongation, when admixed 5 with an otherwise transcriptionally capable or competent composition. The P-TEFb enzyme complex may also be termed "transcriptionally active" or "elongationally active".

The methods for identifying a candidate transcriptional inhibitor may further comprise testing the candidate transcriptional inhibitor so identified in a transcription elongation assay, 10 wherein inhibition of transcription elongation confirms the identification of a transcriptional inhibitor. The transcription elongation assay may be described as comprising the steps of:

- a) preparing a transcriptionally competent composition capable of generating elongated RNA transcripts and comprising effective amounts of DNA template, P-TEFb enzyme complex, RNA polymerase II, each four nucleotides and ATP; and
- b) determining the ability of the transcriptionally competent composition to generate elongated RNA transcripts in the presence of the candidate transcriptional inhibitor and in the absence of the candidate transcriptional inhibitor, wherein a reduction in the amount of elongated RNA or mRNA transcripts in the presence of the candidate is indicative of the identification or confirmation of a transcriptional inhibitor.

25 The invention thus provides transcriptional inhibitors, prepared by a process comprising testing a candidate transcriptional inhibitor substance for the ability to inhibit P-TEFb-mediated phosphorylation of RNA polymerase II and identifying a transcriptional inhibitor as a candidate substance that inhibits the phosphorylation.

The invention also provides methods for identifying an HIV Tat protein, comprising contacting a composition suspected of containing an HIV Tat protein with a human P-TEFb composition under conditions effective to allow the formation of bound protein complexes and detecting the bound protein complexes so formed. The human P-TEFb composition may 5 comprise a human P-TEFb kinase subunit, preferably comprising the amino acid sequence of SEQ ID NO:6. Alternatively, the human P-TEFb composition may comprise a human P-TEFb large subunit, preferably comprising the amino acid sequence of SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50. The human P-TEFb composition may also comprise a human P-TEFb complex comprising a P-TEFb kinase subunit and a P-TEFb large subunit.

10

Given that the studies of the present inventor demonstrate the general importance of human P-TEFb in viral transcription, and show binding to VP16, the invention further provides methods for identifying other transcriptional activator proteins, comprising contacting a composition suspected of containing a transcriptional activator protein with a human P-TEFb composition under conditions effective to allow the formation of bound protein complexes and detecting the bound protein complexes so formed. The detection of a protein that binds to the human P-TEFb composition is indicative of the identification of a transcriptional activator protein.

Also provided in the present invention is a method for identifying a candidate viral transcription inhibitor, comprising testing a candidate substance for the ability to inhibit the binding of a viral transcriptional transactivator protein to a P-TEFb composition comprising a human P-TEFb subunit under effective binding conditions, wherein inhibition of binding is indicative of a positive candidate viral transcription inhibitor. The viral transcriptional transactivator protein may be attached to a solid support, as can the P-TEFb composition. The viral transcriptional transactivator protein may be an HIV Tat protein, or an adenoviral E1A protein or a herpes virus VP16 protein.

30 The P-TEFb composition may comprise a human P-TEFb kinase subunit, preferably comprising the amino acid sequence of SEQ ID NO:6. Alternatively, the P-TEFb composition

may comprise a human P-TEFb large subunit, preferably comprising the amino acid sequence of SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50. In other aspects of the present invention, the P-TEFb composition comprises a human P-TEFb complex comprising a P-TEFb kinase subunit and a P-TEFb large subunit.

5

Such assays are in effect "ELISA type" assays. Accordingly, the effective binding conditions are binding and washing conditions, as may be determined by contacting a composition comprising the viral transcriptional transactivator protein with the P-TEFb composition under conditions effective to allow the formation of bound protein complexes and 10 detecting the bound protein complexes so formed after removing the non-specifically bound protein species. The bound protein complexes may be detected by means of a detectable label that is operatively attached to the viral transcriptional transactivator protein, or alternatively by means of a first anti-viral antibody that binds to the viral transcriptional transactivator protein, preferably wherein the first anti-viral antibody is operatively attached to a detectable label. Also, the bound protein complexes may be detected by means of a detectable label that is operatively 15 attached to the human P-TEFb subunit.

In other aspects, the bound protein complexes may be detected by means of a first anti-P-TEFb antibody that binds to the human P-TEFb subunit, either the human P-TEFb kinase 20 subunit, the human P-TEFb large subunit or an intact human P-TEFb enzyme complex, preferably wherein the first anti-P-TEFb antibody is operatively attached to a detectable label.

In still further aspects of the present invention, the bound protein complexes may be detected by means of a first anti-viral or anti-P-TEFb antibody in combination with a second 25 antibody, the second antibody having binding affinity for the first antibody and being operatively attached to a detectable label. The effective binding conditions may further comprise admixing a human cell nuclear extract with the viral transcriptional transactivator protein and the P-TEFb composition.

The invention provides extended methods of confirming the identification of a candidate viral transcription inhibitor, which further comprise testing the positive candidate viral transcription inhibitor identified in a viral transcription elongation assay, wherein inhibition of viral transcription elongation is indicative of an active candidate viral transcription inhibitor.

5 The viral transcription elongation assay may be described as comprising the steps of:

- 10 a) preparing a transcriptionally competent composition capable of generating elongated viral RNA transcripts and comprising effective amounts of viral nucleic acid template, viral transcriptional transactivator protein, most preferably from the corresponding virus, P-TEFb enzyme complex, RNA polymerase II, each of the four nucleotides and ATP; and
- 15 b) determining the ability of the transcriptionally competent composition to generate elongated viral RNA transcripts in the presence of the positive candidate viral transcription inhibitor and in the absence of the positive candidate inhibitor candidate, wherein a reduction in the amount of elongated viral RNA or viral mRNA transcripts in the presence of the candidate is indicative of the identification or confirmation of an active candidate viral transcriptional inhibitor.

20 In certain aspects, the method may further comprise testing the active candidate viral transcription inhibitor identified in a dual transcription elongation assay, wherein inhibition of viral transcription elongation only in the presence of the viral transcriptional transactivator protein confirms the identification of a viral transcription inhibitor. The dual transcription elongation assay may be described as comprising the steps of:

- 25 a) preparing a first transcriptionally competent composition capable of generating elongated human RNA transcripts, the composition comprising effective amounts of human nucleic acid template, P-TEFb enzyme complex, RNA polymerase II, each of the four nucleotides and ATP;

30

- b) preparing a second transcriptionally competent composition capable of generating elongated viral RNA transcripts, the composition comprising effective amounts of viral nucleic acid template, viral transcriptional transactivator protein, most preferably from the corresponding virus, P-TEFb enzyme complex, RNA polymerase II, each of the four nucleotides and ATP; and
 - c) identifying an active candidate viral transcription inhibitor that inhibits the generation of the elongated viral RNA transcripts by the second transcriptionally competent composition and that does not inhibit the generation of the elongated human RNA transcripts by the first transcriptionally competent composition.

The second transcriptionally competent composition may alternatively comprise any viral template any corresponding viral protein, as exemplified by an adenovirus or a herpes virus nucleic acid template and the transcriptional transactivator proteins, adenovirus E1A or herpes virus VP16.

Although the pre-testing or pre-screening for effective candidate inhibitors, using one or more of the binding assays described above, is believed to be an effective strategy, there is no need that such binding assays be first conducted in order to identify a candidate viral transcription inhibitor. Accordingly, the present invention further provides a method for identifying a candidate viral transcription inhibitor, comprising testing a candidate substance for the ability to inhibit viral RNA elongation in a functional viral transcription elongation assay, wherein inhibition of viral RNA elongation is indicative of an active candidate viral transcription inhibitor.

Preferably, anti-viral screening methods comprise testing the active candidate viral transcription inhibitors in parallel human and viral transcription elongation assays, wherein the presence of the active candidate inhibitor reduces viral, but not human, transcription elongation in the parallel assays.

Thus, the present invention provides viral transcription inhibitors, prepared by a process comprising testing a candidate viral transcription inhibitor substance for the ability to inhibit the binding of a viral transcriptional transactivator protein to a human P-TEFb composition and identifying a viral transcriptional inhibitor as a candidate substance that inhibits the binding under otherwise effective binding conditions. The inhibitors may be dispersed in a pharmaceutically acceptable medium, or admixed with a pharmaceutically acceptable diluent or excipient.

The invention additionally provides an HIV inhibitor, or pharmaceutical formulation thereof, which may be prepared by a process that comprises the steps of:

- a) preparing a first transcriptionally competent composition capable of generating elongated human RNA transcripts, the composition comprising effective amounts of human nucleic acid template, P-TEFb enzyme complex, RNA polymerase II, each of the four nucleotides and ATP;
- b) preparing a second transcriptionally competent composition capable of generating elongated HIV RNA transcripts, the composition comprising effective amounts of HIV nucleic acid template, HIV Tat protein, P-TEFb enzyme complex, RNA polymerase II, each of the four nucleotides and ATP; and
- c) identifying an HIV inhibitor that inhibits the generation of elongated HIV RNA transcripts by the second transcriptionally competent composition but that does not inhibit the generation of elongated human RNA transcripts by the first transcriptionally competent composition.

Also provided by the present invention are methods of inhibiting viral replication, as exemplified by HIV replication, comprising contacting a cell suspected of being infected with a virus, such as HIV, with an amount of the instant inhibitors effective to inhibit viral, or HIV,

RNA elongation in the cell. The cell may be located within an animal, when a therapeutically effective amount of the inhibitor is administered to the animal.

BRIEF DESCRIPTION OF THE DRAWINGS

5

The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

10

FIG. 1. Requirement of the CTD for productive elongation. Quantitation of transcription by RNA polymerase II treated with chymotrypsin for the indicated times. Transcripts were loaded onto a silver stained 6-15% SDS polyacrylamide gel and autoradiographed. Autoradiographs were scanned using a BIO-RAD Model GS-670 Imaging Densitometer. Areas of the silver stained gel corresponding to subunit IIa, indicative of the intact largest polymerase subunit, were quantitated, normalized to the zero (0) digestion time and plotted as CTD remaining. The portions of the autoradiograph indicated as runoff and abortive (520 nucleotide runoff from the actin promoter) transcripts were quantitated and plotted in parallel.

20 FIG. 2A and FIG 2B. Inhibition of productive elongation by DRB and H-8. Transcription reactions in K_cN extract were conducted for 20 min in the presence of the either 0.7 μ M DRB (FIG. 2A) or 7 μ M H-8 (FIG. 2B). Quantitation of levels of runoff transcripts were determined using a Packard InstantImagerTM. Relative levels of runoff at 50% inhibition are shown.

25

FIG. 3. Comparison of the composition and properties of P-TEFb and TFIIH. Quantitation of continuous labeling K_c-FT based transcription assays. Dried silver stained, 6-15% SDS polyacrylamide gels were imaged using the Packard InstantImagerTM and the portions of the gel indicated as runoff were quantitated and normalized to the no addition lanes.

FIG. 4. Inhibition of processive elongation complexes by DRB and H8. The results of a pulse-chase transcription reaction with several DRB and H8 concentrations. The 2 min pulse +/- DRB and H8 was followed by a 45 sec. chase. The reactions were analyzed on a 6%-acrylamide-TBE-urea gel. The quantitation of runoff transcripts from dried polyacrylamide gels were imaged using an Packard InstantImager™ and the portion of the gel indicated as runoff was quantitated and expressed as percent (%) of the runoff compared to the control. The values were normalized by subtracting the drug-insensitive runoff found at the highest drug concentrations used.

10

FIG. 5. Effect of *Drosophila* P-TEFb and DNA titration on transcription. A two min pulse/1 min chase was performed using HeLa nuclear extract and DNA template concentrations of 5, 10, 20, and 40 μ g/ml. Templates were preincubated for 20 min at 30°C in the presence or absence of P-TEFb. The transcripts were analyzed on a 6% acrylamide-TBE-urea gel and quantitation of labeled runoff transcripts, in CPMs, in the presence of DRB was subtracted from the runoff without DRB to obtain DRB-sensitive runoff.

15

FIG. 6. Transcription reactions utilized either control depleted HeLa nuclear extract (cHNE), or PITALRE depleted extract (dHNE). Baculovirus produced recombinant human

20 P-TEFb comprised of PITALRE and cyclin HBL1-1 (rhP-TEFb) was added to reactions as indicated.

25

FIG. 7A and FIG. 7B. DRB and H-8 inhibition of transcription and human P-TEFb activity. Plot of radioactivity in runoff or pol IIo after quantitation using a Packard InstantImager and normalization to the starting amount (100). (FIG. 7A) Transcription of HIV LTR template (633-nt runoff). (FIG. 7B) CTD kinase assay using immunoprecipitated human P-TEFb and *Drosophila* RNA polymerase II as substrate described by Price (Price *et al.*, 1987; Price 1995).

FIG. 8. Quantitation data for *in vitro* Tat transactivation using a continuous labeling protocol with the indicated templates and extracts and similar reaction mixtures which were subjected to a 2 minute pulse and the short transcripts generated were less than 30 nucleotides. Amount of runoff or short transcripts for each template/extract combination was normalized to the corresponding lane with no DRB or Tat added. Runoff transcripts were analyzed in a 6% TBE/Urea gel.

SEQUENCE SUMMARY

10	SEQ ID NO:1	full length cDNA of the <i>Drosophila</i> P-TEFb small (kinase) subunit
	SEQ ID NO:2	amino acid sequence of the <i>Drosophila</i> P-TEFb small subunit
	SEQ ID NO:3	full length cDNA of the <i>Drosophila</i> P-TEFb large subunit
	SEQ ID NO:4	amino acid sequence of the <i>Drosophila</i> P-TEFb large subunit
	SEQ ID NO:5	discovered to be the cDNA sequence of human small subunit
15	SEQ ID NO:6	discovered to be the amino acid sequence of human small subunit
	SEQ ID NO:7	primer for <i>Drosophila</i> RNA polymerase II large subunit
	SEQ ID NO:8	primer for <i>Drosophila</i> RNA polymerase II large subunit
	SEQ ID NO:9	degenerate primer for the <i>Drosophila</i> P-TEFb small (kinase) subunit
20	SEQ ID NO:10	degenerate primer for the <i>Drosophila</i> P-TEFb small (kinase) subunit
	SEQ ID NO:11	primer for the <i>Drosophila</i> P-TEFb small subunit
	SEQ ID NO:12	primer for the <i>Drosophila</i> P-TEFb small subunit
	SEQ ID NO:13	primer for the <i>Drosophila</i> P-TEFb small subunit
	SEQ ID NO:14	primer for the <i>Drosophila</i> P-TEFb small subunit
25	SEQ ID NO:15	primer for the <i>Drosophila</i> P-TEFb small subunit
	SEQ ID NO:16	primer for the <i>Drosophila</i> P-TEFb small subunit
	SEQ ID NO:17	primer for the <i>Drosophila</i> P-TEFb small subunit
	SEQ ID NO:18	primer for the <i>Drosophila</i> P-TEFb small subunit
	SEQ ID NO:19	degenerate primer for the <i>Drosophila</i> P-TEFb large subunit
	SEQ ID NO:20	degenerate primer for the <i>Drosophila</i> P-TEFb large subunit
30	SEQ ID NO:21	degenerate primer for the <i>Drosophila</i> P-TEFb large subunit

5

10

20

25

30

	SEQ ID NO:22	primer for the <i>Drosophila</i> P-TEFb large subunit
	SEQ ID NO:23	primer for the <i>Drosophila</i> P-TEFb large subunit
	SEQ ID NO:24	primer for the <i>Drosophila</i> P-TEFb large subunit
	SEQ ID NO:25	primer for the <i>Drosophila</i> P-TEFb large subunit
5	SEQ ID NO:26	primer for the <i>Drosophila</i> P-TEFb large subunit
	SEQ ID NO:27	primer for the <i>Drosophila</i> P-TEFb large subunit
	SEQ ID NO:28	primer for the <i>Drosophila</i> P-TEFb large subunit
	SEQ ID NO:29	primer for the <i>Drosophila</i> P-TEFb large subunit
	SEQ ID NO:30	primer for the <i>Drosophila</i> P-TEFb large subunit
10	SEQ ID NO:31	peptide sequence for part of the <i>Drosophila</i> P-TEFb small subunit
	SEQ ID NO:32	peptide sequence for part of the <i>Drosophila</i> P-TEFb small subunit
	SEQ ID NO:33	peptide sequence for part of the <i>Drosophila</i> P-TEFb small subunit
	SEQ ID NO:34	peptide sequence for part of the <i>Drosophila</i> P-TEFb small subunit
	SEQ ID NO:35	peptide sequence for part of the <i>Drosophila</i> P-TEFb large subunit
	SEQ ID NO:36	peptide sequence for part of the <i>Drosophila</i> P-TEFb large subunit
	SEQ ID NO:37	peptide sequence for part of the <i>Drosophila</i> P-TEFb large subunit
	SEQ ID NO:38	peptide sequence for part of the <i>Drosophila</i> P-TEFb large subunit
	SEQ ID NO:39	peptide sequence for part of the <i>Drosophila</i> P-TEFb large subunit
	SEQ ID NO:40	peptide sequence for part of the <i>Drosophila</i> P-TEFb large subunit
	SEQ ID NO:41	primer for the human P-TEFb small subunit cDNA
	SEQ ID NO:42	primer for the human P-TEFb small subunit cDNA
	SEQ ID NO:43	full length cDNA of the human P-TEFb large subunit HBL1
	SEQ ID NO:44	coding sequence of human P-TEFb large subunit clone HBL1-1
	SEQ ID NO:45	amino acid sequence of human P-TEFb large subunit clone HBL1-1
25	SEQ ID NO:46	coding sequence of human P-TEFb large subunit clone HBL1-2
	SEQ ID NO:47	amino acid sequence of human P-TEFb large subunit clone HBL1-2
	SEQ ID NO:48	full length cDNA of human P-TEFb large subunit clone HBL3
	SEQ ID NO:49	coding sequence of human P-TEFb large subunit clone HBL3
	SEQ ID NO:50	amino acid sequence of human P-TEFb large subunit clone HBL3

DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

The present invention addresses one or more shortcomings in the prior art through the cloning and further characterization of various subunits of the enzyme, termed positive transcription elongation factor b (P-TEFb), involved in the regulation of RNA elongation. It is herein shown that viral processing, through the joint binding of viral proteins and the carboxyl terminal domain (CTD) of RNA polymerase II and *via* the transfer of phosphate groups to RNA polymerase II, is P-TEFb dependent. The invention relates particularly to the molecular cloning of P-TEFb subunits, including human subunits, to the purification of the recombinant enzyme, to compounds that are capable of inhibiting xenobiotic activators, in particular viral activators, of RNA elongation, and to methods for the identification and use of further inhibitory compounds.

A certain object of the present invention is therefore to provide methods for obtaining P-TEFb enzymes, by purification of the recombinant enzyme from host cells engineered to express the constituent subunits, which methods are proposed to be generally applicable to the purification of all species of P-TEFb.

It is an additional objective of the invention to provide methods for obtaining these enzymes in a relatively purified form, allowing their use in predictive assays for identifying compounds having the ability to reduce the activity of or inhibit viral-driven RNA elongation activity, particularly in the context of retroviruses, such as HIV, and even adenoviruses.

It is a still further object of the invention to identify classes of compounds which demonstrate inhibition of xenobiotic activators of RNA elongation along with a potential application of these compounds in the treatment of diseases connected with human immunodeficiency virus (HIV), herpes simplex virus (HSV) and other viruses and retroviruses.

I. P-TEFb Functions

The claimed invention provides a surprising link between the process of productive elongation and phosphorylation of the CTD of RNA polymerase II. The inventor found that

removal of the CTD by limited proteolysis prohibits the transition into productive elongation. In correlation with this, the inventor found that a factor required for the transition into productive elongation, P-TEFb, is a CTD kinase that, in contrast to earlier candidates, is relevant in intact elongation systems.

5

P-TEFb is an essential component of the positive transcription elongation system which regulates the production of long transcripts *in vitro* (Marshall and Price, 1992). The subunit composition and activity of P-TEFb does not match that of any published *Drosophila* or human transcription factor. In particular, P-TEFb is unlike the elongation factors TFIIIF (factor 5) or DmS-II because it does not stimulate the elongation rate of purified RNA polymerase on dC-tailed templates. P-TEFb was shown to be distinct from *Drosophila* TFIIH by its function in *in vitro* transcription, CTD kinase activity and polypeptide composition. Yankulov *et al.* (1995) proposed that the effect of DRB and H-8 on transcription was due to inhibition of the kinase activity of TFIIH and suggested that the TFIIH associated kinase controlled elongation by RNA polymerase II. The possibility of TFIIH playing a role in elongation control remains, but the present invention directly demonstrates that P-TEFb functions to control elongation and that TFIIH cannot substitute for P-TEFb.

The conclusions of Yankulov *et al.* (1995) were based on the observation that DRB or H-8 inhibited the TFIIH kinase and the appearance of runoff transcripts during transcription with similar dose-response curves. The reported kinase assays were performed at 7.5 μ M ATP and the transcription assays were performed at 500 μ M NTPs. Because different triphosphate conditions were used in the two assays a valid comparison of inhibition curves cannot be made. It is likely that the DRB-sensitivity detected during transcription was due to inhibition of P-TEFb rather than TFIIH, although this conclusion could only be deduced after the discoveries of the present invention.

A slight stimulation of the elongation rate of RNA polymerase II in a dC-tailed template assay was seen with some P-TEFb containing fractions, but the activity did not correlate with

P-TEFb activity. This activity was chromatographically separated from P-TEFb. Based on chromatographic properties, the factor identified by Chodosh *et al.* (1989) in crude fractions from HeLa nuclear extract was not P-TEFb, but possibly the human equivalent of factor 2.

5 Factor 6 was described as a dispensable factor that gave a quantitative stimulation of the amount of runoff transcript seen in reconstructions of transcription initiation (Price *et al.*, 1987). Factor 6 was a DEAE flowthrough fraction from a 0.75 M HGKEDP (buffer comprising 25 mM HEPES, 15% glycerol, gradient concentrations of KCl, 0.1 mM EDTA, 1 mM DTT, 0.1% of a saturated solution of phenylmethylsulfonyl fluoride in isopropanol) step of the phosphocellulose
10 column. The fraction containing factor 6 contained 10 - 15% of the proteins in the nuclear extract and was, therefore, very crude. Another stimulatory activity, factor 7 was also found in the 0.75 M HGKEDP phosphocellulose fraction and this factor bound to the DEAE (Price *et al.*, 1987).

15 Neither of the factors could be further purified at the time because they only had small stimulatory effects and they could not be found during elution of further chromatographic steps. Because both factor 6 and 7 had only small stimulatory effects to RNA elongation, it is unlikely that either factor is P-TEFb, but given that a crude column extract was examined, it is reasonable to surmise that, even though P-TEFb was not detected, it was present in the crude extract as were potentially hundreds of other proteins.
20

25 The particular chromatographic properties of P-TEFb and the subunit composition of P-TEFb distinguish it from any of the known basal initiation factors with the possible exception of TFIIJ (Flores *et al.*, 1992). At least one form of IIJ elutes from P-11 above 0.5 M KCl and the purified factor has two subunits of 33 and 95 kDa. Whereas, P-TEFb elutes from P-11 below 0.5 M KCl and the subunits of P-TEFb are 46.8 kDa and 121 kDa, respectively, as calculated by computer analysis with the genetic analysis program GeneRunner®. A HeLa ss-DNA agarose column fraction that contains both IIA and IIJ does not substitute for P-TEFb.

P-TEFb or a similar protein may play a role in HIV-1 transcription since productive elongation is similar to the stimulation of highly processive elongation complexes by Tat (Marciniak and Sharp, 1991). Recently, Tat-SF was identified as a factor required for Tat activation of HIV transcription (Zhou and Sharp, 1995). The chromatography of this protein on 5 anion exchange resins is most similar to factor 2 or P-TEFa, not P-TEFb. The relation of Tat-SF to P-TEFb is unclear. Recently, a kinase that can associate with the HIV-1 Tat protein was shown to have DRB-sensitive CTD kinase activity (Herrmann and Rice, 1995), but the protein was not purified or characterized.

A specific interaction between HIV-1 and HIV-2 Tat proteins with a HeLa cell protein kinase, termed Tat-associated kinase (TAK), has been demonstrated (Herrmann and Rice, 1993). Tat is a viral protein that acts as a activator of transcription of the HIV viral transcription unit. It binds to short nascent HIV transcripts and causes RNA polymerase II to synthesize mRNA sized transcripts. It has been suggested that a CTD kinase associates with Tat *in vivo*. It has been shown that human cell extracts contain a protein that is able to phosphorylate the CTD of the large subunit of RNA polymerase II (Herrmann and Rice, 1995), but this is a long way from the molecular characterization of "TAK". The TAK activity was inhibited by DRB, which has been shown to selectively inhibit Tat function *in vivo* and *in vitro* (Herrmann and Rice, 1995; Marciniak and Sharp, 1991).

Recent work by Yang *et al.* (1996) has shown that Tat proteins specifically associate with TAK activity *in vivo* and require the CTD of RNA polymerase II for function. An unidentified 42 kDa protein (SEQ ID NO:6) was found to cochromatograph with the TAK activity. They speculated that this protein may be a subunit of TAK. The small subunit of P-TEFb has a similar size to this unidentified protein (Marshall and Price, 1995). Surprisingly, the inventor's comparison of the cDNA sequence of the small subunit of P-TEFb (SEQ ID NO:1) with the cDNA sequence SEQ ID NO:5 of the 42 kDa protein showed an unexpected 71% identity and 82% conserved similarity between the two proteins, suggesting to the inventor that the two proteins are homologous to each other and that P-TEFb is homologous to TAK.

The present invention demonstrates that the human homolog of the small subunit of *Drosophila* P-TEFb binds to a GST-Tat fusion protein, (GST-Tat 1 86R), but not to a GST-Tat transactivation defective protein, (GST-Tat 1 86R P18IS) as described in the Examples herein.

5 The GST-Tat 1 86R/P-TEFb protein complex can phosphorylate the CTD of RNA polymerase II, corroborating the conclusion that TAK is the human homolog of P-TEFb.

Phosphorylation of the CTD of the large subunit of RNA polymerase II normally occurs during the late stages of initiation or early during elongation (Dahmus, 1994; Dahmus, 1995).

10 An intact CTD is required for *Drosophila* elongation complexes to escape early blocks to elongation and enter productive elongation. The serine and threonine rich CTD is a substrate for a large number of kinases (Dahmus, 1994). However, those working in this field were, prior to the present invention, unable to determine which kinase(s) actually phosphorylates the CTD *in vivo* with functional consequences.

PCT/US98/03620
15
X 6/27/98
20

Recently, Tat-SF was proposed as a factor required for Tat activation of HIV transcription (Zhou and Sharp, 1995). Based on the chromatographic analysis of Tat-SF, it appears to be most similar to factor 2 or P-TEFa, two factors which stimulate RNA elongation but do not appear to be essential for RNA elongation to occur. The present invention is therefore surprising in that it clarifies the confusion in the prior art by providing a purified kinase which is essential for the transition into productive RNA elongation. This protein, named P-TEFb, has some homology to cyclin proteins although this similarity could not have been predicted prior to cloning and phosphorylating the CTD of RNA polymerase II.

25 The present invention shows that P-TEFb is distinct from a number of CTD kinases and *in vivo* studies of the present inventor demonstrate that the transition into productive elongation is controlled by P-TEFb-dependent phosphorylation of the CTD. Subunit composition, DRB sensitivity and functional properties distinguish P-TEFb from TFIIH or the TFIIH associated kinase. P-TEFb functions during elongation and is not stably associated with the transcription

30 complex at any time (Kephart *et al.*, 1992; Marshall and Price, 1992; Marshall and Price, 1995).

This further discriminates between P-TEFb and TFIIH as well as the kinase associated with the SRB complex and DNA PK which are part of the initiation complex or otherwise bound to the template. The subunit composition and chromatographic properties of P-TEFb are not similar to any known kinase. Anti-phosphotyrosine antibodies failed to react with RNA polymerase II 5 phosphorylated by P-TEFb, indicating that P-TEFb is a serine/threonine kinase.

The CTD is required for transcription *in vivo* (Zehring *et al.*, 1988; Nonet *et al.*, 1987; Bartolomei *et al.*, 1988; Allison *et al.*, 1988; Brickey and Greenleaf, 1995), but demonstrating a general requirement for the CTD *in vitro* requires the use of crude extracts (Li and Kornberg, 10 1994) or a more purified system which includes specific elongation control factors (Marshall and Price, 1995). However, in more purified systems lacking the elongation control factors some promoters still demonstrate a requirement for the CTD. The CTD is not required for transcription from the *Drosophila* actin 5C promoter in a *Drosophila* system (Zehring *et al.*, 1988) or from the Ad-2 ML promoter in a HeLa system (Kim and Dahmus, 1989). The CTD is required for transcription from the murine DHFR promoter in a HeLa system (Kang and Dahmus, 1993; Thompson *et al.*, 1989; Akoulitchev *et al.*, 1995). Akoulitchev *et al.* (1995), showed that an early step in initiation from the DHFR promoter (formation of the first phosphodiester bond) required the CTD.

A model for the control of elongation has been described which is based on results obtained from a *Drosophila* *in vitro* transcription system (Kephart *et al.*, 1992; Marshall and Price, 1992) and is consistent with data obtained *in vitro* and *in vivo* from many studies. Key features of the model are that all RNA polymerase II molecules that initiate from a promoter are destined to produce only short transcripts in a process termed abortive elongation. Abortive 20 elongation is distinct from abortive initiation because the abortive transcripts are 10 to 20 times longer during abortive elongation and presumably the polymerase in the abortive elongation complexes must relocate the promoter after producing an abortive transcript to bring about reinitiation. Escape from this negative control is accomplished through the action of the positive transcription elongation system which allows productive elongation.

Elongation control can be divided into two temporally differentiated stages during which specific elongation factors function. The transition into productive elongation occurs during the first stage. Negative transcription elongation factors (N-TEF) influence RNA polymerase II molecules that have initiated from a promoter causing them to enter abortive elongation which is
5 characterized by the generation of short transcripts (Kephart *et al.*, 1992; Marshall and Price, 1992). Some of the components of N-TEF seem to be associated with the preinitiation complex (Marshall and Price, 1992), but one component, factor 2, acts after initiation as an RNA polymerase II transcript release factor (Xie and Price, 1996). The generation of highly processive elongation complexes involves several P-TEFs (Marshall and Price, 1995).

10

15

20

The second stage of the elongation control process occurs after RNA polymerase II has entered productive elongation at which point the polymerase may be further affected by other factors. The first factor identified, S-II, suppresses pausing by RNA polymerase II at specific sites (Reinberg and Roeder, 1987; SivaRaman *et al.*, 1990; Sluder *et al.*, 1989) through a transcript cleavage mechanism (Guo and Price, 1993; Izban and Luse, 1993; Reines, 1992). The required initiation factor TFIIF (*Drosophila* factor 5 or RAP 30/74), also stimulates the elongation rate of RNA polymerase II (Bengal *et al.*, 1991; Burton *et al.*, 1988; Flores *et al.*, 1989; Kephart *et al.*, 1994; Price *et al.*, 1989). Other factors, TFIIX (Bengal *et al.*, 1991), S-III (Bradsher *et al.*, 1993a; Bradsher *et al.*, 1993b), and ELL (Shilatifard *et al.*, 1996) stimulate elongation in a manner similar to TFIIF.

25

30

A number of *Drosophila* genes including HSP70 have early blocked polymerases in the IIA form, while downstream elongating polymerases are highly phosphorylated (O'Brien *et al.*, 1994). In injected *Xenopus* oocytes, GAL4 based activators stimulated elongation in a DRB-sensitive manner (Yankulov *et al.*, 1994). It is possible that transcriptional activators work with or through P-TEFb. Dubois *et al.* (1994) found that addition of DRB or H-8 to HeLa cells caused a rapid decrease in the amount of RNA polymerase IIo. Egyhazi *et al.* (1996) recently showed that DRB inhibited the phosphorylation of RNA polymerase II to the IIo form in *Chironomus tentans*. Their results indicated that addition of DRB immediately stopped the incorporation of phosphate into the polymerase, but that polymerases in productive elongation

complexes maintained their hyperphosphorylated states until they completed their current round of transcription.

The substrate specificity of P-TEFb and the kinetics of P-TEFb to phosphorylate CTD were determined under a variety of conditions as described herein in the Examples. The nucleotide requirements of P-TEFb CTD kinase were similar to those reported for HeLa TFIIH (Lu *et al.*, 1992), CTD-K1 (Payne and Dahmus, 1993), and rat liver δ (Serizawa *et al.*, 1992), but were different from those for yeast factor b (Feaver *et al.*, 1991) and CTD-K2 (Payne and Dahmus, 1993). The kinetics of phosphorylation using different concentrations of kinase is not consistent with processive P-TEFb action during the phosphorylation of purified RNA polymerase II. P-TEFb does prefer to phosphorylate a CTD that has already been phosphorylated which may have implications for its function after initiation.

II. P-TEFb Genes and Expression

A. DNA Segments

Important aspects of the present invention concern isolated DNA segments and recombinant vectors encoding P-TEFb, and the creation and use of recombinant host cells through the application of DNA technology, that express P-TEFb. DNA segments, recombinant vectors, recombinant host cells and expression methods using sequences of the P-TEFb small subunit and large subunit are also provided.

Each of the foregoing genes are included within all aspects of the following description. The present invention concerns DNA segments, isolatable from insect, mammalian and human cells, that are free from total genomic DNA and that are capable of expressing a P-TEFb protein or polypeptide subunit that has kinase activity, or that has similarity to cyclin proteins, both of which subunits are essential for productive RNA elongation. Such proteins will herein be termed "P-TEFb proteins".

As used herein, the term "DNA segment" refers to a DNA molecule that has been isolated free of total genomic DNA of a particular species. Therefore, a DNA segment encoding a P-TEFb subunit refers to a DNA segment that contains P-TEFb coding sequences yet is isolated away from, or purified free from, total mammalian or human genomic DNA. Included within the 5 term "DNA segment", are DNA segments and smaller fragments of such segments, and also recombinant vectors, including, for example, plasmids, cosmids, phage, viruses, and the like.

Similarly, a DNA segment comprising an isolated or purified P-TEFb kinase or large 10 subunit protein gene refers to a DNA segment including P-TEFb kinase or large subunit protein coding sequences and, in certain aspects, regulatory sequences, isolated substantially away from other naturally occurring genes or protein encoding sequences. In this respect, the term "gene" is used for simplicity to refer to a functional protein, polypeptide or peptide encoding unit. As will be understood by those in the art, this functional term includes both genomic sequences, complementary DNA (cDNA) sequences and smaller engineered gene segments that express, or may be adapted to express, P-TEFb proteins, polypeptides, domains, peptides, fusion proteins and mutants.

"Isolated substantially away from other coding sequences" means that the gene of interest, in this case a P-TEFb kinase or large subunit protein gene, forms the significant part of the coding region of the DNA segment, and that the DNA segment does not contain large portions of naturally-occurring coding DNA, such as large chromosomal fragments or other functional genes or cDNA coding regions. Of course, this refers to the DNA segment as originally isolated, and does not exclude genes or coding regions later added to the segment by the hand of man.

25

In particular embodiments, as described in detail in the foregoing 'Summary' section, the invention concerns isolated DNA segments and recombinant vectors that encode substantially full length P-TEFb kinase subunit proteins or polypeptide that include a contiguous amino acid sequence of at least about 24 amino acids, and more preferably, of at least about, 25, 27, 30, 35, 30 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 200 amino acids or so from SEQ ID NO:2, or a

biologically functional equivalent thereof. Even more preferably, the genes encode a protein with a sequence essentially as set forth in SEQ ID NO:2.

Substantially full length P-TEFb kinase subunit genes preferably include a contiguous nucleic acid sequence of at least about 722 nucleotides, and more preferably, of at least about 725, 750, 800, 825, 850, 900 or so nucleotides from between position 115 and position 1326 or 1327 of SEQ ID NO:1, or a biologically functional equivalent thereof. More preferably, the isolated genes and DNA segments have a nucleic acid sequence essentially as set forth in SEQ ID NO:1.

10

The substantially full length P-TEFb large subunit genes of the invention generally encode a protein or polypeptide that includes a contiguous amino acid sequence of at least about 6 or 7 amino acids or so, or more preferably, of at least about 8, 10, 12, 14, 16, 18, 20, 22 or 25 amino acids or so from SEQ ID NO:4, SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50, or a biologically functional equivalent thereof; or the genes and DNA segments hybridize to such a coding sequence under stringent hybridization conditions.

More preferably, the substantially full length P-TEFb large subunit genes encode a P-TEFb large subunit protein that includes a contiguous amino acid sequence of at least about 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 200 amino acids or so from SEQ ID NO:4, SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50, or a biologically functional equivalent thereof; or hybridize to such a coding sequence under stringent hybridization conditions. Most preferably, the substantially full length P-TEFb large subunit genes encode a P-TEFb large subunit having a sequence essentially as set forth in the amino acid sequence of SEQ ID NO:4, SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50, or a biologically functional equivalent thereof; or hybridize to such a coding sequence under stringent hybridization conditions.

In certain embodiments, the substantially full length P-TEFb large subunit genes include a contiguous nucleic acid sequence of at least about 120, 150 or 200 or so nucleotides from 30 between position 716 and position 4006 of SEQ ID NO:3, or include a contiguous nucleic acid

sequence of at least about 360, 400, 450 or 500 or so nucleotides from a coding region from SEQ ID NO:43 or SEQ ID NO:48 (*e.g.*, SEQ ID NO:44, SEQ ID NO:46 or SEQ ID NO:49); or a biologically functionally equivalent thereof; or will hybridize to such a coding sequence under stringent hybridization conditions.

5

Most preferably, the substantially full length P-TEFb large subunit genes comprise an isolated coding region having a nucleic acid sequence essentially as set forth in SEQ ID NO:3, SEQ ID NO:44, SEQ ID NO:46 or SEQ ID NO:49, or a biologically functional equivalent thereof; or will hybridize to such a coding sequence under stringent hybridization conditions.

10

The term "a sequence essentially as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50" means that the sequence substantially corresponds to a portion of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50 and has relatively few amino acids that are not identical to, or a biologically functional equivalent of, the amino acids of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50. The term "biologically functional equivalent" is well understood in the art and is further defined in detail herein. Accordingly, sequences that have between about 85% and about 90%; or more preferably, between about 91% and about 95%; or even more preferably, between about 96% and about 99%; of amino acids that are identical or functionally equivalent to the amino acids of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50 will be sequences that are "essentially as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50", provided the biological activity of the protein is maintained.

15

20

25

30

In certain other embodiments, the invention concerns isolated DNA segments and recombinant vectors that include within their sequence a nucleic acid sequence essentially as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:43, SEQ ID NO:46 or SEQ ID NO:48. The term "essentially as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:43, SEQ ID NO:46 or SEQ ID NO:48" is used in the same sense as described above and means that the nucleic acid sequence substantially corresponds to a portion of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:43,

SEQ ID NO:46 or SEQ ID NO:48 and has relatively few codons that are not identical, or functionally equivalent, to the codons of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:43, SEQ ID NO:46 or SEQ ID NO:48. Again, DNA segments that encode proteins exhibiting kinase activity or large functional subunits will be most preferred.

5

The term "functionally equivalent codon" is used herein to refer to codons that encode the same amino acid, such as the six codons for arginine or serine, and also refers to codons that encode biologically equivalent amino acids (see Codon Table, below).

10

CODON TABLE

Amino Acids			Codons
Alanine	Ala	A	GCA GCC GCG GCU
Cysteine	Cys	C	UGC UGU
Aspartic acid	Asp	D	GAC GAU
Glutamic acid	Glu	E	GAA GAG
Phenylalanine	Phe	F	UUC UUU
Glycine	Gly	G	GGA GGC GGG GGU
Histidine	His	H	CAC CAU
Isoleucine	Ile	I	AUA AUC AUU
Lysine	Lys	K	AAA AAG
Leucine	Leu	L	UUA UUG CUA CUC CUG CUU
Methionine	Met	M	AUG
Asparagine	Asn	N	AAC AAU
Proline	Pro	P	CCA CCC CCG CCU
Glutamine	Gln	Q	CAA CAG
Arginine	Arg	R	AGA AGG CGA CGC CGG CGU
Serine	Ser	S	AGC AGU UCA UCC UCG UCU
Threonine	Thr	T	ACA ACC ACG ACU
Valine	Val	V	GUU GUC GUG GUU
Tryptophan	Trp	W	UGG
Tyrosine	Tyr	Y	UAC UAU

It will also be understood that amino acid and nucleic acid sequences may include additional residues, such as additional N- or C-terminal amino acids or 5' or 3' sequences, and yet still be essentially as set forth in one of the sequences disclosed herein, so long as the

15

sequence meets the criteria set forth above, including the maintenance of biological protein activity where protein expression is concerned. The addition of terminal sequences particularly applies to nucleic acid sequences that may, for example, include various non-coding sequences flanking either of the 5' or 3' portions of the coding region or may include various internal sequences, *i.e.*, introns, which are known to occur within genes.

Excepting intronic or flanking regions, and allowing for the degeneracy of the genetic code, sequences that have between about 75% and about 79%; or more preferably, between about 80% and about 89%; or even more preferably, between about 90% and about 99%; of nucleotides that are identical to the nucleotides of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:43, SEQ ID NO:46 or SEQ ID NO:48 will be sequences that are "essentially as set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:43, SEQ ID NO:46 or SEQ ID NO:48".

Sequences that are essentially the same as those set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:43, SEQ ID NO:46 or SEQ ID NO:48 may also be functionally defined as sequences that are capable of hybridizing to a nucleic acid segment containing the complement of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:43, SEQ ID NO:46 or SEQ ID NO:48 under relatively stringent conditions. Suitable relatively stringent hybridization conditions will be well known to those of skill in the art.

Suitable standard hybridization conditions for the present invention include, for example, hybridization in 50% formamide, 5× Denhardts' solution, 5× SSC, 25 mM sodium phosphate, 0.1% SDS and 100 µg/ml of denatured salmon sperm DNA at 42°C for 16 h followed by 1h sequential washes with 0.1× SSC, 0.1% SDS solution at 60°C to remove the desired amount of background signal. Lower stringency hybridization conditions for the present invention include, for example, hybridization in 35% formamide, 5× Denhardts' solution, 5× SSC, 25 mM sodium phosphate, 0.1% SDS and 100 µg/ml denatured salmon sperm DNA or *E. coli* DNA at 42°C for 16 h followed by sequential washes with 0.8× SSC, 0.1% SDS at 55°C. Those of skill in the art will recognize that conditions can be readily adjusted to obtain the desired level of stringency.

Naturally, the present invention also encompasses DNA segments that are complementary, or essentially complementary, to the sequence set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:43, SEQ ID NO:46 or SEQ ID NO:48. Nucleic acid sequences that are 5 "complementary" are those that are capable of base-pairing according to the standard Watson-Crick complementarity rules. As used herein, the term "complementary sequences" means nucleic acid sequences that are substantially complementary, as may be assessed by the same nucleotide comparison set forth above, or as defined as being capable of hybridizing to the nucleic acid segment of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:43, SEQ ID NO:46 or SEQ 10 ID NO:48 under relatively stringent conditions such as those described immediately above.

The nucleic acid segments of the present invention, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol. 15

For example, nucleic acid fragments may be prepared from any region of the presently disclosed sequences that include contiguous sequences characterized as follows: about 117-125 or about 126-135 nucleotides of SEQ ID NO:3; about 315-320 or about 321-330 nucleotides of SEQ ID NO:48; about 360-370 or about 371-380 nucleotides of either SEQ ID NO:43 or SEQ ID NO:46. The sequences may be up to about 30,000 or 20,000, or about 10,000, or about 5,000 20 base pairs in length, with segments of about 3,000 being preferred in certain cases.

DNA segments with total lengths of about 1,000, about 500, about 300, about 200, and about 150 base pairs in length (including all intermediate lengths) are also contemplated to be useful for certain sequences as described above. For the smaller fragments, probes and primers 30 the sequences will generally be selected from the following defined regions: between positions

1-258, 320-345 and 1244-1457 of SEQ ID NO:1; positions 587-964, 1156-1711, 1764-3287, 3460-3775 and 3800-4328 of SEQ ID NO:3; and preferably positions, 1-244, 297-546, 867-1142, 1895-2331, 2821-2890, 3341-3442, 3953-3860 and 4491-4528 of SEQ ID NO:43; and positions 1-209, 418-667, 919-1031, 2045-2164 and 2219-2360 of SEQ ID NO:48.

5

It will be readily understood that "intermediate lengths", in these contexts, means any length between the quoted ranges, bearing in mind that the stated constraints on lengths and regions of complementarity. For example, 19, 20, 21, *etc.*; 24, 25, 26, *etc.*; 29, 30, 31, *etc.*; 48, 49, 50, 51, *etc.*; 75, 76, 77, 78, 79, 80 *etc.*; 100, 101, 102, 103 *etc.*; 118, 119, 120, 121 *etc.*; 127, 128, 129, 130, 131, *etc.*; 316, 317, 318, 319, *etc.*; 322, 323, 324, 325, 326, *etc.*; 361, 362, 363, 364, *etc.*; 372, 373, 374, 375, *etc.*; including all integers through the 400-500; 500-1,000; 1,000-2,000; 2,000-3,000; 3,000-5,000; 5,000-10,000 ranges, up to and including sequences of about 12,001, 12,002, 13,001, 13,002, 15,000, 20,000, 30,000 and the like as is appropriate for each sequence.

10
15

The various probes and primers designed around the disclosed nucleotide sequences of the present invention may be of any length. By assigning numeric values to a sequence, for example, the first residue is 1, the second residue is 2, *etc.*, an algorithm defining all primers can be proposed:

20
25

n to n + y

where n is an integer from 1 to the last number of the sequence and y is the length of the primer minus one, where n + y does not exceed the last number of the sequence. Thus, for a 125-mer, the probes correspond to bases 1 to 125, 2 to 126, 3 to 127 ... and so on. For a 315-mer, the probes correspond to bases 1 to 315, 2 to 316, 3 to 317 ... and so on. For a 360-mer, the probes correspond to bases 1 to 360, 2 to 361, 3 to 362 ... and so on.

30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995
1000
1005
1010
1015
1020
1025
1030
1035
1040
1045
1050
1055
1060
1065
1070
1075
1080
1085
1090
1095
1100
1105
1110
1115
1120
1125
1130
1135
1140
1145
1150
1155
1160
1165
1170
1175
1180
1185
1190
1195
1200
1205
1210
1215
1220
1225
1230
1235
1240
1245
1250
1255
1260
1265
1270
1275
1280
1285
1290
1295
1300
1305
1310
1315
1320
1325
1330
1335
1340
1345
1350
1355
1360
1365
1370
1375
1380
1385
1390
1395
1400
1405
1410
1415
1420
1425
1430
1435
1440
1445
1450
1455
1460
1465
1470
1475
1480
1485
1490
1495
1500
1505
1510
1515
1520
1525
1530
1535
1540
1545
1550
1555
1560
1565
1570
1575
1580
1585
1590
1595
1600
1605
1610
1615
1620
1625
1630
1635
1640
1645
1650
1655
1660
1665
1670
1675
1680
1685
1690
1695
1700
1705
1710
1715
1720
1725
1730
1735
1740
1745
1750
1755
1760
1765
1770
1775
1780
1785
1790
1795
1800
1805
1810
1815
1820
1825
1830
1835
1840
1845
1850
1855
1860
1865
1870
1875
1880
1885
1890
1895
1900
1905
1910
1915
1920
1925
1930
1935
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025
2030
2035
2040
2045
2050
2055
2060
2065
2070
2075
2080
2085
2090
2095
2100
2105
2110
2115
2120
2125
2130
2135
2140
2145
2150
2155
2160
2165
2170
2175
2180
2185
2190
2195
2200
2205
2210
2215
2220
2225
2230
2235
2240
2245
2250
2255
2260
2265
2270
2275
2280
2285
2290
2295
2300
2305
2310
2315
2320
2325
2330
2335
2340
2345
2350
2355
2360
2365
2370
2375
2380
2385
2390
2395
2400
2405
2410
2415
2420
2425
2430
2435
2440
2445
2450
2455
2460
2465
2470
2475
2480
2485
2490
2495
2500
2505
2510
2515
2520
2525
2530
2535
2540
2545
2550
2555
2560
2565
2570
2575
2580
2585
2590
2595
2600
2605
2610
2615
2620
2625
2630
2635
2640
2645
2650
2655
2660
2665
2670
2675
2680
2685
2690
2695
2700
2705
2710
2715
2720
2725
2730
2735
2740
2745
2750
2755
2760
2765
2770
2775
2780
2785
2790
2795
2800
2805
2810
2815
2820
2825
2830
2835
2840
2845
2850
2855
2860
2865
2870
2875
2880
2885
2890
2895
2900
2905
2910
2915
2920
2925
2930
2935
2940
2945
2950
2955
2960
2965
2970
2975
2980
2985
2990
2995
3000
3005
3010
3015
3020
3025
3030
3035
3040
3045
3050
3055
3060
3065
3070
3075
3080
3085
3090
3095
3100
3105
3110
3115
3120
3125
3130
3135
3140
3145
3150
3155
3160
3165
3170
3175
3180
3185
3190
3195
3200
3205
3210
3215
3220
3225
3230
3235
3240
3245
3250
3255
3260
3265
3270
3275
3280
3285
3290
3295
3300
3305
3310
3315
3320
3325
3330
3335
3340
3345
3350
3355
3360
3365
3370
3375
3380
3385
3390
3395
3400
3405
3410
3415
3420
3425
3430
3435
3440
3445
3450
3455
3460
3465
3470
3475
3480
3485
3490
3495
3500
3505
3510
3515
3520
3525
3530
3535
3540
3545
3550
3555
3560
3565
3570
3575
3580
3585
3590
3595
3600
3605
3610
3615
3620
3625
3630
3635
3640
3645
3650
3655
3660
3665
3670
3675
3680
3685
3690
3695
3700
3705
3710
3715
3720
3725
3730
3735
3740
3745
3750
3755
3760
3765
3770
3775
3780
3785
3790
3795
3800
3805
3810
3815
3820
3825
3830
3835
3840
3845
3850
3855
3860
3865
3870
3875
3880
3885
3890
3895
3900
3905
3910
3915
3920
3925
3930
3935
3940
3945
3950
3955
3960
3965
3970
3975
3980
3985
3990
3995
4000
4005
4010
4015
4020
4025
4030
4035
4040
4045
4050
4055
4060
4065
4070
4075
4080
4085
4090
4095
4100
4105
4110
4115
4120
4125
4130
4135
4140
4145
4150
4155
4160
4165
4170
4175
4180
4185
4190
4195
4200
4205
4210
4215
4220
4225
4230
4235
4240
4245
4250
4255
4260
4265
4270
4275
4280
4285
4290
4295
4300
4305
4310
4315
4320
4325
4330
4335
4340
4345
4350
4355
4360
4365
4370
4375
4380
4385
4390
4395
4400
4405
4410
4415
4420
4425
4430
4435
4440
4445
4450
4455
4460
4465
4470
4475
4480
4485
4490
4495
4500
4505
4510
4515
4520
4525
4530
4535
4540
4545
4550
4555
4560
4565
4570
4575
4580
4585
4590
4595
4600
4605
4610
4615
4620
4625
4630
4635
4640
4645
4650
4655
4660
4665
4670
4675
4680
4685
4690
4695
4700
4705
4710
4715
4720
4725
4730
4735
4740
4745
4750
4755
4760
4765
4770
4775
4780
4785
4790
4795
4800
4805
4810
4815
4820
4825
4830
4835
4840
4845
4850
4855
4860
4865
4870
4875
4880
4885
4890
4895
4900
4905
4910
4915
4920
4925
4930
4935
4940
4945
4950
4955
4960
4965
4970
4975
4980
4985
4990
4995
5000
5005
5010
5015
5020
5025
5030
5035
5040
5045
5050
5055
5060
5065
5070
5075
5080
5085
5090
5095
5100
5105
5110
5115
5120
5125
5130
5135
5140
5145
5150
5155
5160
5165
5170
5175
5180
5185
5190
5195
5200
5205
5210
5215
5220
5225
5230
5235
5240
5245
5250
5255
5260
5265
5270
5275
5280
5285
5290
5295
5300
5305
5310
5315
5320
5325
5330
5335
5340
5345
5350
5355
5360
5365
5370
5375
5380
5385
5390
5395
5400
5405
5410
5415
5420
5425
5430
5435
5440
5445
5450
5455
5460
5465
5470
5475
5480
5485
5490
5495
5500
5505
5510
5515
5520
5525
5530
5535
5540
5545
5550
5555
5560
5565
5570
5575
5580
5585
5590
5595
5600
5605
5610
5615
5620
5625
5630
5635
5640
5645
5650
5655
5660
5665
5670
5675
5680
5685
5690
5695
5700
5705
5710
5715
5720
5725
5730
5735
5740
5745
5750
5755
5760
5765
5770
5775
5780
5785
5790
5795
5800
5805
5810
5815
5820
5825
5830
5835
5840
5845
5850
5855
5860
5865
5870
5875
5880
5885
5890
5895
5900
5905
5910
5915
5920
5925
5930
5935
5940
5945
5950
5955
5960
5965
5970
5975
5980
5985
5990
5995
6000
6005
6010
6015
6020
6025
6030
6035
6040
6045
6050
6055
6060
6065
6070
6075
6080
6085
6090
6095
6100
6105
6110
6115
6120
6125
6130
6135
6140
6145
6150
6155
6160
6165
6170
6175
6180
6185
6190
6195
6200
6205
6210
6215
6220
6225
6230
6235
6240
6245
6250
6255
6260
6265
6270
6275
6280
6285
6290
6295
6300
6305
6310
6315
6320
6325
6330
6335
6340
6345
6350
6355
6360
6365
6370
6375
6380
6385
6390
6395
6400
6405
6410
6415
6420
6425
6430
6435
6440
6445
6450
6455
6460
6465
6470
6475
6480
6485
6490
6495
6500
6505
6510
6515
6520
6525
6530
6535
6540
6545
6550
6555
6560
6565
6570
6575
6580
6585
6590
6595
6600
6605
6610
6615
6620
6625
6630
6635
6640
6645
6650
6655
6660
6665
6670
6675
6680
6685
6690
6695
6700
6705
6710
6715
6720
6725
6730
6735
6740
6745
6750
6755
6760
6765
6770
6775
6780
6785
6790
6795
6800
6805
6810
6815
6820
6825
6830
6835
6840
6845
6850
6855
6860
6865
6870
6875
6880
6885
6890
6895
6900
6905
6910
6915
6920
6925
6930
6935
6940
6945
6950
6955
6960
6965
6970
6975
6980
6985
6990
6995
7000
7005
7010
7015
7020
7025
7030
7035
7040
7045
7050
7055
7060
7065
7070
7075
7080
7085
7090
7095
7100
7105
7110
7115
7120
7125
7130
7135
7140
7145
7150
7155
7160
7165
7170
7175
7180
7185
7190
7195
7200
7205
7210
7215
7220
7225
7230
7235
7240
7245
7250
7255
7260
7265
7270
7275
7280
7285
7290
7295
7300
7305
7310
7315
7320
7325
7330
7335
7340
7345
7350
7355
7360
7365
7370
7375
7380
7385
7390
7395
7400
7405
7410
7415
7420
7425
7430
7435
7440
7445
7450
7455
7460
7465
7470
7475
7480
7485
7490
7495
7500
7505
7510
7515
7520
7525
7530
7535
7540
7545
7550
7555
7560
7565
7570
7575
7580
7585
7590
7595
7600
7605
7610
7615
7620
7625
7630
7635
7640
7645
7650
7655
7660
7665
7670
7675
7680
7685
7690
7695
7700
7705
7710
7715
7720
7725
7730
7735
7740
7745
7750
7755
7760
7765
7770
7775
7780
7785
7790
7795
7800
7805
7810
7815
7820
7825
7830
7835
7840
7845
7850
7855
7860
7865
7870
7875
7880
7885
7890
7895
7900
7905
7910
7915
7920
7925
7930
7935
7940
7945
7950
7955
7960
7965
7970
7975
7980
7985
7990
7995
8000
8005
8010
8015
8020
8025
8030
8035
8040
8045
8050
8055
8060
8065
8070
8075
8080
8085
8090
8095
8100
8105
8110
8115
8120
8125
8130
8135
8140
8145
8150
8155
8160
8165
8170
8175
8180
8185
8190
8195
8200
8205
8210
8215
8220
8225
8230
8235
8240
8245
8250
8255
8260
8265
8270
8275
8280
8285
8290
8295
8300
8305
8310
8315
8320
8325
8330
8335
8340
8345
8350
8355
8360
8365
8370
8375
8380
8385
8390
8395
8400
8405
8410
8415
8420
8425
8430
8435
8440
8445
8450
8455
8460
8465
8470
8475
8480
8485
8490
8495
8500
8505
8510
8515
8520
8525
8530
8535
8540
8545
8550
8555
8560
8565
8570
8575
8580
8585
8590
8595
8600
8605
8610
8615
8620
8625
8630
8635<br

include these coding regions themselves, coding regions bearing selected alterations or modifications in the basic coding region, or they may encode larger polypeptides that nevertheless include such coding regions or may encode biologically functional equivalent proteins or peptides that have variant amino acids sequences.

5

The DNA segments of the present invention encompass biologically functional equivalent P-TEFb and kinase proteins and peptides. Such sequences may arise as a consequence of codon redundancy and functional equivalency that are known to occur naturally within nucleic acid sequences and the proteins thus encoded. Alternatively, functionally equivalent proteins or peptides may be created *via* the application of recombinant DNA technology, in which changes in the protein structure may be engineered, based on considerations of the properties of the amino acids being exchanged. Changes designed by man may be introduced through the application of site-directed mutagenesis techniques, *e.g.*, to introduce improvements to the antigenicity of the protein or to test mutants in order to examine transcription, elongation or Tat binding activity at the molecular level.

One may also prepare fusion proteins and peptides, *e.g.*, where the P-TEFb or kinase protein coding regions are aligned within the same expression unit with other proteins or peptides having desired functions, such as for purification or immunodetection purposes (*e.g.*, proteins that may be purified by affinity chromatography and enzyme label coding regions, respectively).

Encompassed by the invention are DNA segments encoding relatively small peptides, such as, for example, peptides of from about 7 to about 50 amino acids in length, and more preferably, of from about 10 to about 30 amino acids in length; and also larger polypeptides up to and including proteins corresponding to the full-length sequences set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:45, SEQ ID NO:47 and SEQ ID NO:50.

The DNA segments of the present invention may be employed for a variety of applications. For example, a particularly useful application concerns the recombinant production

of the individual subunits or proteins or peptides whose structure is derived from that of the subunits, or in the recombinant production of the holoenzyme following co-expression of the two subunits. Additionally, the P-TEFb -encoding DNA segments of the present invention can also be used in the preparation of nucleic acid probes or primers, which can, for example, be used in
5 the identification and cloning of P-TEFb genes or related genomic sequences, or in the study of subunit(s) expression, and the like.

B. Nucleic Acid Detection

In addition to their use in directing the expression of the PTEF kinase or large subunit
10 protein, the nucleic acid sequences disclosed herein also have a variety of other uses. For example, they also have utility as probes or primers in nucleic acid hybridization embodiments.

1. Hybridization

The use of a hybridization probe, *e.g.*, of about 14-20, 25-30, 50, 75, 100, 120, 150, 200,
15 250, 300 or 360 or so nucleotides in length allows the formation of a duplex molecule that is both stable and selective. Molecules having complementary sequences over stretches greater than 20 bases in length are generally preferred, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of particular hybrid molecules obtained. In using sequences from any region of those disclosed herein, one may generally prefer to design nucleic acid molecules having stretches of 120 to 360 nucleotides, or even longer where desired. By choosing from more unique regions, as already disclosed herein, those of skill in the art will appreciate that smaller probess and primers may be designed and utilized, *e.g.*, of between about
20 14-20, 25, 30, 35 or so nucleotides in length.

25 All fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means or by introducing selected sequences into recombinant vectors for recombinant production. These chemical means can include PCR™ technology of U.S. Patent 4,603,102 (herein incorporated by reference) or by introducing selected sequences into recombinant vectors for recombinant production.

Accordingly, the nucleotide sequences of the invention may be used for their ability to selectively form duplex molecules with complementary stretches of genes or RNAs or to provide primers for amplification of DNA or RNA from tissues. Depending on the application envisioned,
5 one will desire to employ varying conditions of hybridization to achieve varying degrees of selectivity of probe towards target sequence.

For certain applications, for example, substitution of nucleotides by site-directed mutagenesis, it is appreciated that lower stringency conditions are required. Under these
10 conditions, hybridization may occur even though the sequences of probe and target strand are not perfectly complementary, but are mismatched at one or more positions. Conditions may be rendered less stringent by increasing salt concentration and decreasing temperature. For example, low stringency hybridization conditions for the present invention provide hybridization in 35% formamide, 5 \times Denhardts' solution, 5 \times SSC, 25 mM sodium phosphate, 0.1% SDS and 100 μ g/ml denatured salmon sperm DNA or *E. coli* DNA at 42°C for 16 hours followed by sequential washes with 0.8 \times SSC, 0.1% SDS at 55°C and allows for cross-species hybridization to homologous proteins to occur. For example, under these conditions the *Drosophila* P-TEFb small subunit cDNA (SEQ ID NO:1) cross hybridizes to the human 42 kDa protein of SEQ ID NO:6. Thus, hybridization conditions can be readily manipulated depending on the desired results.
15 Alternative hybridization conditions which are useful are given in Examples 3 and 4. Of course, the hybridization conditions chosen are dependent upon the objective that is to be achieved.
20

In other embodiments, more stringent hybridization may be achieved under conditions of, for example, 50% formamide, 5 \times Denhardts' solution, 5 \times SSC, 25 mM sodium phosphate, 0.1%
25 SDS and 100 μ g/ml of denatured salmon sperm DNA at 42°C for 16 h followed by 1hour sequential washes with 0.1 \times SSC, 0.1% SDS solution at 60°C to remove the desired amount of background signal.

In certain embodiments, it will be advantageous to employ nucleic acid sequences of the present invention in combination with an appropriate means, such as a label, for determining hybridization. A wide variety of appropriate indicator means are known in the art, including fluorescent, radioactive, enzymatic or other ligands, such as avidin/biotin, which are capable of being detected. In preferred embodiments, one may desire to employ a fluorescent label or an enzyme tag such as urease, alkaline phosphatase or peroxidase, instead of radioactive or other environmentally undesirable reagents. In the case of enzyme tags, colorimetric indicator substrates are known that can be employed to provide a detection means visible to the human eye or spectrophotometrically, to identify specific hybridization with complementary nucleic acid-containing samples.

In general, it is envisioned that the hybridization probes described herein will be useful both as reagents in solution hybridization, as in PCR™, for detection of expression of corresponding genes, as well as in embodiments employing a solid phase. In embodiments involving a solid phase, the test DNA (or RNA) is adsorbed or otherwise affixed to a selected matrix or surface. This fixed, single-stranded nucleic acid is then subjected to hybridization with selected probes under desired conditions. The selected conditions will depend on the particular circumstances based on the particular criteria required (depending, for example, on the G+C content, type of target nucleic acid, source of nucleic acid, size of hybridization probe, *etc.*). Following washing of the hybridized surface to remove non-specifically bound probe molecules, hybridization is detected, or even quantified, by means of the label.

2. Amplification and PCR™

Nucleic acid used as a template for amplification is isolated from cells contained in the biological sample, according to standard methodologies (Sambrook *et al.*, 1989). The nucleic acid may be genomic DNA or fractionated or whole cell RNA. Where RNA is used, it may be desired to convert the RNA to a cDNA.

Pairs of primers that selectively hybridize to nucleic acids corresponding to P-TEFb, kinase protein or a mutant thereof are contacted with the isolated nucleic acid under conditions that permit selective hybridization. The term "primer", as defined herein, is meant to encompass any nucleic acid that is capable of priming the synthesis of a nascent nucleic acid in a template-dependent process. Typically, primers are oligonucleotides from ten to twenty base pairs in length, but longer sequences can be employed. Primers may be provided in double-stranded or single-stranded form, although the single-stranded form is preferred.

Once hybridized, the nucleic acid:primer complex is contacted with one or more enzymes that facilitate template-dependent nucleic acid synthesis. Multiple rounds of amplification, also referred to as "cycles," are conducted until a sufficient amount of amplification product is produced.

Next, the amplification product is detected. In certain applications, the detection may be performed by visual means. Alternatively, the detection may involve indirect identification of the product *via* chemiluminescence, radioactive scintigraphy of incorporated radiolabel or fluorescent label or even *via* a system using electrical or thermal impulse signals (Affymax technology).

A number of template dependent processes are available to amplify the marker sequences present in a given template sample. One of the best known amplification methods is the polymerase chain reaction (referred to as PCRTM) which is described in detail in U.S. Patent Nos. 4,683,195, 4,683,202 and 4,800,159, and each incorporated herein by reference in entirety.

Briefly, in PCRTM, two primer sequences are prepared that are complementary to regions on opposite complementary strands of the marker sequence. An excess of deoxynucleoside triphosphates are added to a reaction mixture along with a DNA polymerase, *e.g.*, *Taq* polymerase. If the marker sequence is present in a sample, the primers will bind to the marker and the polymerase will cause the primers to be extended along the marker sequence by adding

on nucleotides. By raising and lowering the temperature of the reaction mixture, the extended primers will dissociate from the marker to form reaction products, excess primers will bind to the marker and to the reaction products and the process is repeated.

5 A reverse transcriptase PCR™ (RT-PCR™) amplification procedure may be performed in order to quantify the amount of mRNA amplified or to prepare cDNA from the desired mRNA. Methods of reverse transcribing RNA into cDNA are well known and described in Sambrook *et al.*, 1989. Alternative methods for reverse transcription utilize thermostable, RNA-dependent DNA polymerases. These methods are described in WO 90/07641, filed December 21, 1990,
10 incorporated herein by reference. Polymerase chain reaction methodologies are well known in the art.

Another method for amplification is the ligase chain reaction ("LCR"), disclosed in EPA No. 320 308, incorporated herein by reference in its entirety. In LCR, two complementary probe pairs are prepared, and in the presence of the target sequence, each pair will bind to opposite complementary strands of the target such that they abut. In the presence of a ligase, the two probe pairs will link to form a single unit. By temperature cycling, as in PCR™, bound ligated units dissociate from the target and then serve as "target sequences" for ligation of excess probe pairs. U.S. Patent 4,883,750 describes a method similar to LCR for binding probe pairs to a target sequence.
15
20

Qbeta Replicase, described in PCT Application No. PCT/US87/00880, incorporated herein by reference, may also be used as still another amplification method in the present invention. In this method, a replicative sequence of RNA that has a region complementary to
25 that of a target is added to a sample in the presence of an RNA polymerase. The polymerase will copy the replicative sequence that can then be detected.

An isothermal amplification method, in which restriction endonucleases and ligases are used to achieve the amplification of target molecules that contain nucleotide 5'-[alpha-thio]-

triposphates in one strand of a restriction site may also be useful in the amplification of nucleic acids in the present invention.

Strand Displacement Amplification (SDA) is another method of carrying out isothermal amplification of nucleic acids which involves multiple rounds of strand displacement and synthesis, *i.e.*, nick translation. A similar method, called Repair Chain Reaction (RCR), involves annealing several probes throughout a region targeted for amplification, followed by a repair reaction in which only two of the four bases are present. The other two bases can be added as biotinylated derivatives for easy detection. A similar approach is used in SDA. Target specific sequences can also be detected using a cyclic probe reaction (CPR). In CPR, a probe having 3' and 5' sequences of non-specific DNA and a middle sequence of specific RNA is hybridized to DNA that is present in a sample. Upon hybridization, the reaction is treated with RNase H, and the products of the probe identified as distinctive products that are released after digestion. The original template is annealed to another cycling probe and the reaction is repeated.

Still another amplification methods described in GB Application No. 2 202 328, and in PCT Application No. PCT/US89/01025, each of which is incorporated herein by reference in its entirety, may be used in accordance with the present invention. In the former application, "modified" primers are used in a PCR™-like, template- and enzyme-dependent synthesis. The primers may be modified by labeling with a capture moiety (*e.g.*, biotin) and/or a detector moiety (*e.g.*, enzyme). In the latter application, an excess of labeled probes are added to a sample. In the presence of the target sequence, the probe binds and is cleaved catalytically. After cleavage, the target sequence is released intact to be bound by excess probe. Cleavage of the labeled probe signals the presence of the target sequence.

Other nucleic acid amplification procedures include transcription-based amplification systems (TAS), including nucleic acid sequence based amplification (NASBA) and 3SR Gingeras *et al.*, PCT Application WO 88/10315, incorporated herein by reference. In NASBA, the nucleic acids can be prepared for amplification by standard phenol/chloroform extraction,

heat denaturation of a clinical sample, treatment with lysis buffer and minispin columns for isolation of DNA and RNA or guanidinium chloride extraction of RNA. These amplification techniques involve annealing a primer which has target specific sequences. Following polymerization, DNA/RNA hybrids are digested with RNase H while double stranded DNA molecules are heat denatured again. In either case the single stranded DNA is made fully double stranded by addition of second target specific primer, followed by polymerization. The double-stranded DNA molecules are then multiply transcribed by an RNA polymerase such as T7 or SP6. In an isothermal cyclic reaction, the RNA's are reverse transcribed into single stranded DNA, which is then converted to double stranded DNA, and then transcribed once again with an RNA polymerase such as T7 or SP6. The resulting products, whether truncated or complete, indicate target specific sequences.

Davey *et al.*, EPA No. 329 822 (incorporated herein by reference in its entirety) disclose a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA), which may be used in accordance with the present invention. The ssRNA is a template for a first primer oligonucleotide, which is elongated by reverse transcriptase (RNA-dependent DNA polymerase). The RNA is then removed from the resulting DNA:RNA duplex by the action of ribonuclease H (RNase H, an RNase specific for RNA in duplex with either DNA or RNA). The resultant ssDNA is a template for a second primer, which also includes the sequences of an RNA polymerase promoter (exemplified by T7 RNA polymerase) 5' to its homology to the template. This primer is then extended by DNA polymerase (exemplified by the large "Klenow" fragment of *E. coli* DNA polymerase I), resulting in a double-stranded DNA ("dsDNA") molecule, having a sequence identical to that of the original RNA between the primers and having additionally, at one end, a promoter sequence. This promoter sequence can be used by the appropriate RNA polymerase to make many RNA copies of the DNA. These copies can then re-enter the cycle leading to very swift amplification. With proper choice of enzymes, this amplification can be done isothermally without addition of enzymes at each cycle. Because of the cyclical nature of this process, the starting sequence can be chosen to be in the form of either DNA or RNA.

Miller *et al.*, PCT Application WO 89/06700 (incorporated herein by reference in its entirety) disclose a nucleic acid sequence amplification scheme based on the hybridization of a promoter/primer sequence to a target single-stranded DNA ("ssDNA") followed by transcription of many RNA copies of the sequence. This scheme is not cyclic, *i.e.*, new templates are not produced from the resultant RNA transcripts. Other amplification methods include "RACE" and "one-sided PCR" (Frohman, 1990, incorporated by reference).

Methods based on ligation of two (or more) oligonucleotides in the presence of nucleic acid having the sequence of the resulting "di-oligonucleotide", thereby amplifying the di-oligonucleotide, may also be used in the amplification step of the present invention.

Following any amplification, it may be desirable to separate the amplification product from the template and the excess primer for the purpose of determining whether specific amplification has occurred. In one embodiment, amplification products are separated by agarose, agarose-acrylamide or polyacrylamide gel electrophoresis using standard methods. See Sambrook *et al.*, 1989.

Alternatively, chromatographic techniques may be employed to effect separation. There are many kinds of chromatography which may be used in the present invention: adsorption, partition, ion-exchange and molecular sieve, and many specialized techniques for using them including column, paper, thin-layer and gas chromatography.

Amplification products must be visualized in order to confirm amplification of the marker sequences. One typical visualization method involves staining of a gel with ethidium bromide and visualization under UV light. Alternatively, if the amplification products are integrally labeled with radio- or fluorometrically-labeled nucleotides, the amplification products can then be exposed to x-ray film or visualized under the appropriate stimulating spectra, following separation.

In one embodiment, visualization is achieved indirectly. Following separation of amplification products, a labeled, nucleic acid probe is brought into contact with the amplified marker sequence. The probe preferably is conjugated to a chromophore but may be radiolabeled. In another embodiment, the probe is conjugated to a binding partner, such as an antibody or 5 biotin, and the other member of the binding pair carries a detectable moiety.

In one embodiment, detection is by Southern blotting and hybridization with a labeled probe. The techniques involved in Southern blotting are well known to those of skill in the art and can be found in many standard books on molecular protocols. See Sambrook *et al.*, 1989. 10 Briefly, amplification products are separated by gel electrophoresis. The gel is then contacted with a membrane, such as nitrocellulose, permitting transfer of the nucleic acid and non-covalent binding. Subsequently, the membrane is incubated with a chromophore-conjugated probe that is capable of hybridizing with a target amplification product. Detection is by exposure of the membrane to x-ray film or ion-emitting detection devices. 15

One example of the foregoing is described in U.S. Patent No. 5,279,721, incorporated by reference herein, which discloses an apparatus and method for the automated electrophoresis and transfer of nucleic acids. The apparatus permits electrophoresis and blotting without external manipulation of the gel and is ideally suited to carrying out methods according to the present invention. 20

All the essential materials and reagents required for detecting P-TEFb or kinase protein markers in a biological sample may be assembled together in a kit. This generally will comprise preselected primers for specific markers. Also included may be enzymes suitable for amplifying 25 nucleic acids including various polymerases (RT, Taq, *etc.*), deoxynucleotides and buffers to provide the necessary reaction mixture for amplification.

Such kits generally will comprise, in suitable means, distinct containers for each individual reagent and enzyme as well as for each marker primer pair. Preferred pairs of primers 30 for amplifying nucleic acids are selected to amplify the sequences specified in SEQ ID NO:1 or

SEQ ID NO:3 or SEQ ID NO:43 or SEQ ID NO:46 or SEQ ID NO:48 such that, for example, nucleic acid fragments are prepared that include a contiguous stretch of nucleotides identical to for example about 14-20, 25, 30, 35, *etc.*; 48, 49, 50, 51, *etc.*; 75, 76, 77, 78, 79, 80 *etc.*; 100, 101, 102, 103 *etc.*; 118, 119, 120, 121 *etc.*; 127, 128, 129, 130, 131, *etc.*; 316, 317, 318, 319, 5 *etc.*; 322, 323, 324, 325, 326, *etc.*; 361, 362, 363, 364, *etc.*; 372, 373, 374, 375, *etc.* of SEQ ID NO:3, SEQ ID NO:48, SEQ ID NO:43 or SEQ ID NO:4, so long as the selected contiguous stretches are from spatially distinct regions. Similar fragments may be prepared which are identical or complimentary to SEQ ID NO:1 such that the fragments do not hybridize to SEQ ID NO:5.

10

15

In another embodiment, such kits will comprise hybridization probes specific for P-TEFb large or kinase proteins chosen from a group including nucleic acids corresponding to the sequences specified in SEQ ID NO:1 or SEQ ID NO:3 or SEQ ID NO:43 or SEQ ID NO:46 or SEQ ID NO:48 or to intermediate lengths of the sequences specified. Such kits generally will comprise, in suitable means, distinct containers for each individual reagent and enzyme as well as for each marker hybridization probe.

PATENT
PENDING
15
20
25
30

3. Other Assays

Other methods for genetic screening to accurately detect genetic changes which may be caused by disease, such as cancers, viral or parasitic infections that alter normal cellular production and processing, in genomic DNA, cDNA or RNA samples may be employed, depending on the specific situation.

For example, one method of screening for genetic variation is based on RNase cleavage 25 of base pair mismatches in RNA/DNA and RNA/RNA heteroduplexes. As used herein, the term "mismatch" is defined as a region of one or more unpaired or mispaired nucleotides in a double-stranded RNA/RNA, RNA/DNA or DNA/DNA molecule. This definition thus includes mismatches due to insertion/deletion mutations, as well as single and multiple base point mutations.

30

U.S. Patent No. 4,946,773 describes an RNase A mismatch cleavage assay that involves annealing single-stranded DNA or RNA test samples to an RNA probe, and subsequent treatment of the nucleic acid duplexes with RNase A. After the RNase cleavage reaction, the RNase is inactivated by proteolytic digestion and organic extraction, and the cleavage products are
5 denatured by heating and analyzed by electrophoresis on denaturing polyacrylamide gels. For the detection of mismatches, the single-stranded products of the RNase A treatment, electrophoretically separated according to size, are compared to similarly treated control duplexes. Samples containing smaller fragments (cleavage products) not seen in the control duplex are scored as +.

10

Currently available RNase mismatch cleavage assays, including those performed according to U.S. Patent No. 4,946,773, require the use of radiolabeled RNA probes. Myers and Maniatis in U.S. Patent No. 4,946,773 describe the detection of base pair mismatches using RNase A. Other investigators have described the use of *E. coli* enzyme, RNase I, in mismatch assays. Because it has broader cleavage specificity than RNase A, RNase I would be a desirable enzyme to employ in the detection of base pair mismatches if components can be found to decrease the extent of non-specific cleavage and increase the frequency of cleavage of mismatches. The use of RNase I for mismatch detection is described in literature from Promega Biotech. Promega markets a kit containing RNase I that is shown in their literature to cleave three out of four known mismatches, provided the enzyme level is sufficiently high.
15
20

The RNase protection assay was first used to detect and map the ends of specific mRNA targets in solution. The assay relies on being able to easily generate high specific activity radiolabeled RNA probes complementary to the mRNA of interest by *in vitro* transcription.
25 Originally, the templates for *in vitro* transcription were recombinant plasmids containing bacteriophage promoters. The probes are mixed with total cellular RNA samples to permit hybridization to their complementary targets, then the mixture is treated with RNase to degrade excess unhybridized probe. Also, as originally intended, the RNase used is specific for single-stranded RNA, so that hybridized double-stranded probe is protected from degradation. After

inactivation and removal of the RNase, the protected probe (which is proportional in amount to the amount of target mRNA that was present) is recovered and analyzed on a polyacrylamide gel.

The RNase Protection assay was adapted for detection of single base mutations. In this type of RNase A mismatch cleavage assay, radiolabeled RNA probes transcribed *in vitro* from wild type sequences, are hybridized to complementary target regions derived from test samples. The test target generally comprises DNA (either genomic DNA or DNA amplified by cloning in plasmids or by PCRTM), although RNA targets (endogenous mRNA) have occasionally been used. If single nucleotide (or greater) sequence differences occur between the hybridized probe and target, the resulting disruption in Watson-Crick hydrogen bonding at that position ("mismatch") can be recognized and cleaved in some cases by single-strand specific ribonuclease. To date, RNase A has been used almost exclusively for cleavage of single-base mismatches, although RNase I has recently been shown as useful also for mismatch cleavage. There are recent descriptions of using the MutS protein and other DNA-repair enzymes for detection of single-base mismatches.

C. Mutagenesis

Site-specific mutagenesis is a technique useful in the preparation of individual peptides, or biologically functional equivalent proteins or peptides, through specific mutagenesis of the underlying DNA. The technique further provides a ready ability to prepare and test sequence variants, incorporating one or more of the foregoing considerations, by introducing one or more nucleotide sequence changes into the DNA. Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Typically, a primer of about 17 to 25 nucleotides in length is preferred, with about 5 to 10 residues on both sides of the junction of the sequence being altered.

In general, the technique of site-specific mutagenesis is well known in the art. As will be appreciated, the technique typically employs a bacteriophage vector that exists in both a single stranded and double stranded form. Typical vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage vectors are commercially available and their use is generally well known to those skilled in the art. Double stranded plasmids are also routinely employed in site directed mutagenesis, which eliminates the step of transferring the gene of interest from a phage to a plasmid.

In general, site-directed mutagenesis is performed by first obtaining a single-stranded vector, or melting of two strands of a double stranded vector which includes within its sequence a DNA sequence encoding the desired protein. An oligonucleotide primer bearing the desired mutated sequence is synthetically prepared. This primer is then annealed with the single-stranded DNA preparation, and subjected to DNA polymerizing enzymes such as *E. coli* polymerase I Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed wherein one strand encodes the original non-mutated sequence and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate cells, such as *E. coli* cells, and clones are selected that include recombinant vectors bearing the mutated sequence arrangement.

The preparation of sequence variants of the selected gene using site-directed mutagenesis is provided as a means of producing potentially useful species and is not meant to be limiting, as there are other ways in which sequence variants of genes may be obtained. For example, recombinant vectors encoding the desired gene may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence variants.

D. Recombinant Vectors, Host Cells and Expression
Recombinant vectors form important further aspects of the present invention. The term "expression vector or construct" means any type of genetic construct containing a nucleic acid coding for a gene product in which part or all of the nucleic acid encoding sequence is capable of being transcribed. The transcript may be translated into a protein, but it need not be. Thus, in

certain embodiments, expression includes both transcription of a gene and translation of a RNA into a gene product. In other embodiments, expression only includes transcription of the nucleic acid, for example, to generate antisense constructs.

5 Particularly useful vectors are contemplated to be those vectors in which the coding portion of the DNA segment, whether encoding a full length protein or smaller peptide, is positioned under the transcriptional control of a promoter. A "promoter" refers to a DNA sequence recognized by the synthetic machinery of the cell, or introduced synthetic machinery, required to initiate the specific transcription of a gene. The phrases "operatively positioned",
10 "under control" or "under transcriptional control" means that the promoter is in the correct location and orientation in relation to the nucleic acid to control RNA polymerase initiation and expression of the gene.

15 The promoter may be in the form of the promoter that is naturally associated with a P-TEFb or kinase protein gene, as may be obtained by isolating the 5' non-coding sequences located upstream of the coding segment or exon, for example, using recombinant cloning and/or PCR™ technology, in connection with the compositions disclosed herein (PCR™ technology is disclosed in U.S. Patent 4,683,202 and U.S. Patent 4,682,195, each incorporated herein by reference).

20 In other embodiments, it is contemplated that certain advantages will be gained by positioning the coding DNA segment under the control of a recombinant, or heterologous, promoter. As used herein, a recombinant or heterologous promoter is intended to refer to a promoter that is not normally associated with a P-TEFb or a kinase protein gene in its natural environment. Such promoters may include promoters normally associated with other genes, and/or promoters isolated from any other bacterial, viral, eukaryotic, or mammalian cell.
25

Naturally, it will be important to employ a promoter that effectively directs the expression of the DNA segment in the cell type, organism, or even animal, chosen for

expression. The use of promoter and cell type combinations for protein expression is generally known to those of skill in the art of molecular biology, for example, see Sambrook *et al.* (1989), incorporated herein by reference. The promoters employed may be constitutive, or inducible, and can be used under the appropriate conditions to direct high level expression of the introduced
5 DNA segment, such as is advantageous in the large-scale production of recombinant proteins or peptides.

At least one module in a promoter functions to position the start site for RNA synthesis. The best known example of this is the TATA box, but in some promoters lacking a TATA box,
10 such as the promoter for the mammalian terminal deoxynucleotidyl transferase gene and the promoter for the SV40 late genes, a discrete element overlying the start site itself helps to fix the place of initiation.

Additional promoter elements regulate the frequency of transcriptional initiation.
15 Typically, these are located in the region 30-110 bp upstream of the start site, although a number of promoters have been shown to contain functional elements downstream of the start site as well. The spacing between promoter elements frequently is flexible, so that promoter function is preserved when elements are inverted or moved relative to one another. In the thymidine kinase (tk) promoter, the spacing between promoter elements can be increased to 50 bp apart before activity begins to decline. Depending on the promoter, it appears that individual elements can function either co-operatively or independently to activate transcription.
20

The particular promoter that is employed to control the expression of a nucleic acid is not believed to be critical, so long as it is capable of expressing the nucleic acid in the targeted cell.
25 Thus, where a human cell is targeted, it is preferable to position the nucleic acid coding region adjacent to and under the control of a promoter that is capable of being expressed in a human cell. Generally speaking, such a promoter might include either a human or viral promoter. Preferred promoters include those derived from HSV, including the HNF1 α promoter. Another preferred embodiment is the tetracycline controlled promoter.

In various other embodiments, the human cytomegalovirus (CMV) immediate early gene promoter, the Simian virus 40 (SV40) early promoter and the Rous sarcoma virus long terminal repeat can be used to obtain high-level expression of transgenes. The use of other viral or 5 mammalian cellular or bacterial phage promoters which are well-known in the art to achieve expression of a transgene is contemplated as well, provided that the levels of expression are sufficient for a given purpose. The following tables list several elements/promoters which may be employed, in the context of the present invention, to regulate the expression of P-TEFb or a kinase protein gene. This list is not intended to be exhaustive of all the possible elements 10 involved in the promotion of transgene expression but, merely, to be exemplary thereof.

Enhancers were originally detected as genetic elements that increased transcription from a promoter located at a distant position on the same molecule of DNA. This ability to act over a large distance had little precedent in classic studies of prokaryotic transcriptional regulation. Subsequent work showed that regions of DNA with enhancer activity are organized much like 15 promoters. That is, they are composed of many individual elements, each of which binds to one or more transcriptional proteins.

The basic distinction between enhancers and promoters is operational. An enhancer region as a whole must be able to stimulate transcription at a distance; this need not be true of a promoter region or its component elements. On the other hand, a promoter must have one or 20 more elements that direct initiation of RNA synthesis at a particular site and in a particular orientation, whereas enhancers lack these specificities. Promoters and enhancers are often overlapping and contiguous, often seeming to have a very similar modular organization.

25

Additionally any promoter/enhancer combination (as per the Eukaryotic Promoter Data Base EPDB) could also be used to drive expression of a transgene. Use of a T3, T7 or SP6 cytoplasmic expression system is another possible embodiment. Eukaryotic cells can support cytoplasmic transcription from certain bacterial promoters if the appropriate bacterial polymerase

is provided, either as part of the delivery complex or as an additional genetic expression construct.

PROMOTER TABLE

PROMOTER	REFERENCES
Immunoglobulin Heavy Chain	Hanerji <i>et al.</i> , 1983; Gilles <i>et al.</i> , 1983; Grosschedl and Baltimore, 1985; Atchinson and Perry, 1986, 1987; Imler <i>et al.</i> , 1987; Weinberger <i>et al.</i> , 1988; Kiledjian <i>et al.</i> , 1988; Porton <i>et al.</i> , 1990
Immunoglobulin Light Chain	Queen and Baltimore, 1983; Picard and Schaffner, 1984
T-Cell Receptor	Luria <i>et al.</i> , 1987, Winoto and Baltimore, 1989; Redondo <i>et al.</i> , 1990
HLA DQ α and DQ β	Sullivan and Peterlin, 1987
β -Interferon	Goodbourn <i>et al.</i> , 1986; Fujita <i>et al.</i> , 1987; Goodbourn and Maniatis, 1985
Interleukin-2	Greene <i>et al.</i> , 1989
Interleukin-2 Receptor	Greene <i>et al.</i> , 1989; Lin <i>et al.</i> , 1990
MHC Class II 5	Koch <i>et al.</i> , 1989
MHC Class II HLA-DRA	Sherman <i>et al.</i> , 1989
β -Actin	Kawamoto <i>et al.</i> , 1988; Ng <i>et al.</i> , 1989
Muscle Creatine Kinase	Jaynes <i>et al.</i> , 1988; Horlick and Benfield, 1989; Johnson <i>et al.</i> , 1989a
Prealbumin (Transthyretin)	Costa <i>et al.</i> , 1988
Elastase I	Omitz <i>et al.</i> , 1987
Metallothionein	Karin <i>et al.</i> , 1987; Culotta and Hamer, 1989
Collagenase	Pinkert <i>et al.</i> , 1987; Angel <i>et al.</i> , 1987
Albumin Gene	Pinkert <i>et al.</i> , 1987, Tronche <i>et al.</i> , 1989, 1990
α -Fetoprotein	Godbout <i>et al.</i> , 1988; Campere and Tilghman, 1989
τ -Globin	Bodine and Ley, 1987; Perez-Stable and Constantini, 1990
β -Globin	Trudel and Constantini, 1987
e-fos	Cohen <i>et al.</i> , 1987
c-HA-ras	Triesman, 1986; Deschamps <i>et al.</i> , 1985
Insulin	Edlund <i>et al.</i> , 1985
Neural Cell Adhesion Molecule (NCAM)	Hirsch <i>et al.</i> , 1990

PROMOTER	REFERENCES
α -1-Antitrypsin	Latimer <i>et al.</i> , 1990
H2B (TH2B) Histone	Hwang <i>et al.</i> , 1990
Mouse or Type I Collagen	Ripe <i>et al.</i> , 1989
Glucose-Regulated Proteins (GRP94 and GRP78)	Chang <i>et al.</i> , 1989
Rat Growth Hormone	Larsen <i>et al.</i> , 1986
Human Serum Amyloid A (SAA)	Edbrooke <i>et al.</i> , 1989
Troponin I (TN I)	Yutzey <i>et al.</i> , 1989
Platelet-Derived Growth Factor	Pech <i>et al.</i> , 1989
Duchenne Muscular Dystrophy	Klamut <i>et al.</i> , 1990
SV40	Banerji <i>et al.</i> , 1981; Moreau <i>et al.</i> , 1981; Sleigh and Lockett, 1985; Firak and Subramanian, 1986; Herr and Clarke, 1986; Imbra and Karin, 1986; Kadesch and Berg, 1986; Wang and Calame, 1986; Ondek <i>et al.</i> , 1987; Kuhl <i>et al.</i> , 1987 Schaffner <i>et al.</i> , 1988
Polyoma	Swartzendruber and Lehman, 1975; Vasseur <i>et al.</i> , 1980; Katinka <i>et al.</i> , 1980, 1981; Tyndell <i>et al.</i> , 1981; Dandolo <i>et al.</i> , 1983; deVilliers <i>et al.</i> , 1984; Hen <i>et al.</i> , 1986; Satake <i>et al.</i> , 1988; Campbell and Villarreal, 1988
Retroviruses	Kriegler and Botchan, 1982, 1983; Levinson <i>et al.</i> , 1982; Kriegler <i>et al.</i> , 1983, 1984a,b, 1988; Bosze <i>et al.</i> , 1986; Miksicek <i>et al.</i> , 1986; Celander and Haseltine, 1987; Thiesen <i>et al.</i> , 1988; Celander <i>et al.</i> , 1988; Chol <i>et al.</i> , 1988; Reisman and Rotter, 1989
Papilloma Virus	Campo <i>et al.</i> , 1983; Lusky <i>et al.</i> , 1983; Spandidos and Wilkie, 1983; Spalholz <i>et al.</i> , 1985; Lusky and Botchan, 1986; Cripe <i>et al.</i> , 1987; Gloss <i>et al.</i> , 1987; Hirochika <i>et al.</i> , 1987, Stephens and Hentschel, 1987; Glue <i>et al.</i> , 1988
Hepatitis B Virus	Bulla and Siddiqui, 1986; Jameel and Siddiqui, 1986; Shaul and Ben-Levy, 1987; Spandau and Lee, 1988; Vannice and Levinson, 1988
Human Immunodeficiency Virus	Muesing <i>et al.</i> , 1987; Hauber and Cullan, 1988; Jakobovits <i>et al.</i> , 1988; Feng and Holland, 1988; Takebe <i>et al.</i> , 1988; Rowen <i>et al.</i> , 1988; Berkhout <i>et al.</i> , 1989; Laspia <i>et al.</i> , 1989; Sharp and Marciniak, 1989; Braddock <i>et al.</i> , 1989
Cytomegalovirus	Weber <i>et al.</i> , 1984; Boshart <i>et al.</i> , 1985; Foecking and Hofstetter, 1986
Gibbon Ape Leukemia Virus	Holbrook <i>et al.</i> , 1987; Quinn <i>et al.</i> , 1989

ENHANCER TABLE

	Inducer	References
MT II	Phorbol Ester (TFA) Heavy metals	Palmiter <i>et al.</i> , 1982; Haslinger and Karin, 1985; Searle <i>et al.</i> , 1985; Stuart <i>et al.</i> , 1985; Imagawa <i>et al.</i> , 1987; Karin ®, 1987; Angel <i>et al.</i> , 1987b; McNeall <i>et al.</i> , 1989
MMTV (mouse mammary tumor virus)	Glucocorticoids	Huang <i>et al.</i> , 1981; Lee <i>et al.</i> , 1981; Majors and Varmus, 1983; Chandler <i>et al.</i> , 1983; Lee <i>et al.</i> , 1984; Fonta <i>et al.</i> , 1985; Sakai <i>et al.</i> , 1986
β-Interferon	poly(rI)X poly(rc)	Tavernier <i>et al.</i> , 1983
Adenovirus 5 E2	Ela	Imperiale and Nevins, 1984
Collagenase	Phorbol Ester (TPA)	Angle <i>et al.</i> , 1987a
Stromelysin	Phorbol Ester (TPA)	Angle <i>et al.</i> , 1987b
SV40	Phorbol Ester (TFA)	Angel <i>et al.</i> , 1987b
Murine MX Gene	Interferon, Newcastle Disease Virus	
GRP78 Gene	A23187	Resendez <i>et al.</i> , 1988
α-2-Macroglobulin	IL-6	Kunz <i>et al.</i> , 1989
Vimentin	Serum	Rittling <i>et al.</i> , 1989
MHC Class I Gene H-2kb	Interferon	Blanar <i>et al.</i> , 1989
HSP70	Ela, SV40 Large T Antigen	Taylor <i>et al.</i> , 1989; Taylor and Kingston, 1990a,b
Proliferin	Phorbol Ester-TPA	Mordacq and Linzer, 1989
Tumor Necrosis Factor	FMA	Hensel <i>et al.</i> , 1989
Thyroid Stimulating Hormone α Gene	Thyroid Hormone	Chatterjee <i>et al.</i> , 1989

Turning to the expression of the P-TEFb or kinase proteins of the present invention, once a suitable clone or clones have been obtained, whether they be cDNA based or genomic, one may proceed to prepare an expression system. The engineering of DNA segment(s) for expression in a prokaryotic or eukaryotic system may be performed by techniques generally known to those of skill in recombinant expression. It is believed that virtually any expression system may be employed in the expression of the proteins of the present invention.

Two *in vitro* systems that best represent the features of the *in vivo* situation are nuclear extracts from HeLa and *Drosophila* K_c cells (K_cN). In HeLa nuclear extracts transcription initiating at the HIV LTR promoter gives rise to an abundance of transcripts terminated close to the promoter and only very limited amounts of longer transcripts. Addition of HIV Tat protein increases the number of DRB-sensitive elongation complexes that are able to make long transcripts (Marciniak and Sharp, 1991). Using *Drosophila* K_cN extracts only a fraction of RNA polymerase II molecules that initiate generate transcripts longer than several hundred nucleotides (Kephart *et al.*, 1992). The addition of DRB to KcN extracts selectively inhibits the production of elongation complexes capable of sustained elongation. Two factors P-TEF (positive transcription elongation factor) and N-TEF (negative transcription elongation factor) were proposed to control this behavior (Marshall and Price, 1992).

The inventor has characterized processive elongation complexes in HeLa nuclear extracts that are either DRB-sensitive or -insensitive. The DRB-sensitive processive elongation complexes produce runoff transcripts at a faster rate than DRB-insensitive processive elongation complexes. The immobilized DNA template studies suggest the existence of a HeLa P-TEFb-like activity. The number of DRB-sensitive processive elongation complexes appears to be determined by the amount of P-TEFb activity. The inventor shows that *Drosophila* and HeLa P-TEFb activities are DRB-sensitive and can act cross-species in the transition into productive elongation.

Both cDNA and genomic sequences are suitable for eukaryotic expression, as the host cell will generally process the genomic transcripts to yield functional mRNA for translation into

protein. Generally speaking, it may be more convenient to employ as the recombinant gene a cDNA version of the gene. It is believed that the use of a cDNA version will provide advantages in that the size of the gene will generally be much smaller and more readily employed to transfect the targeted cell than will a genomic gene, which will typically be up to an order of magnitude larger than the cDNA gene. However, the inventor does not exclude the possibility of employing a genomic version of a particular gene where desired.

In expression, one will typically include a polyadenylation signal to effect proper polyadenylation of the transcript. The nature of the polyadenylation signal is not believed to be crucial to the successful practice of the invention, and any such sequence may be employed. Preferred embodiments include the SV40 polyadenylation signal and the bovine growth hormone polyadenylation signal, convenient and known to function well in various target cells. Also contemplated as an element of the expression cassette is a terminator. These elements can serve to enhance message levels and to minimize read through from the cassette into other sequences.

A specific initiation signal also may be required for efficient translation of coding sequences. These signals include the ATG initiation codon and adjacent sequences. Exogenous translational control signals, including the ATG initiation codon, may need to be provided. One of ordinary skill in the art would readily be capable of determining this and providing the necessary signals. It is well known that the initiation codon must be "in-frame" with the reading frame of the desired coding sequence to ensure translation of the entire insert. The exogenous translational control signals and initiation codons can be either natural or synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements.

It is proposed that the small subunit of P-TEFb may be co-expressed with the large subunit of P-TEFb, wherein the proteins may be co-expressed in the same cell or wherein one subunit, for example the small subunit, may be provided to a cell that already has the other subunit, for example the large subunit, of P-TEFb. Co-expression may be achieved by co-transfected the cell with two distinct recombinant vectors, each bearing a copy of either the

respective DNA. Alternatively, a single recombinant vector may be constructed to include the coding regions for both of the subunits, which could then be expressed in cells transfected with the single vector. In either event, the term "co-expression" herein refers to the expression of both of the subunits of P-TEFb in the same recombinant cell.

5

As used herein, the terms "engineered" and "recombinant" cells are intended to refer to a cell into which an exogenous DNA segment or gene, such as a cDNA or gene encoding a P-TEFb or kinase protein has been introduced. Therefore, engineered cells are distinguishable from naturally occurring cells which do not contain a recombinantly introduced exogenous DNA
10 segment or gene. Engineered cells are thus cells having a gene or genes introduced through the hand of man. Recombinant cells include those having an introduced cDNA or genomic gene, and also include genes positioned adjacent to a promoter not naturally associated with the particular introduced gene.

15 To express a recombinant P-TEFb or kinase protein, whether mutant or wild-type, in accordance with the present invention one would prepare an expression vector that comprises a P-TEFb- or kinase protein-encoding nucleic acid under the control of one or more promoters. To bring a coding sequence "under the control of" a promoter, one positions the 5' end of the transcription initiation site of the transcriptional reading frame generally between about 1 and about 50 nucleotides "downstream" of (*i.e.*, 3' of) the chosen promoter. The "upstream" 20 promoter stimulates transcription of the DNA and promotes expression of the encoded recombinant protein. This is the meaning of "recombinant expression" in this context.

25 Many standard techniques are available to construct expression vectors containing the appropriate nucleic acids and transcriptional/translational control sequences in order to achieve protein or peptide expression in a variety of host-expression systems. Cell types available for expression include, but are not limited to, bacteria, such as *E. coli* and *B. subtilis* transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors.

Certain examples of prokaryotic hosts are *E. coli* strain RR1, *E. coli* LE392, *E. coli* B, *E. coli* X 1776 (ATCC No. 31537) as well as *E. coli* W3110 (F-, lambda-, prototrophic, ATCC No. 273325); bacilli such as *Bacillus subtilis*; and other enterobacteriaceae such as *Salmonella typhimurium*, *Serratia marcescens*, and various *Pseudomonas* species.

5

In general, plasmid vectors containing replicon and control sequences which are derived from species compatible with the host cell are used in connection with these hosts. The vector ordinarily carries a replication site, as well as marking sequences which are capable of providing phenotypic selection in transformed cells. For example, *E. coli* is often transformed using pBR322, a plasmid derived from an *E. coli* species. pBR322 contains genes for ampicillin and tetracycline resistance and thus provides easy means for identifying transformed cells. The pBR plasmid, or other microbial plasmid or phage must also contain, or be modified to contain, promoters which can be used by the microbial organism for expression of its own proteins.

10
15 In addition, phage vectors containing replicon and control sequences that are compatible with the host microorganism can be used as transforming vectors in connection with these hosts. For example, the phage lambda GEMTM-11 may be utilized in making a recombinant phage vector which can be used to transform host cells, such as *E. coli* LE392.

20 Further useful vectors include pIN vectors and pGEX vectors, for use in generating glutathione S-transferase (GST) soluble fusion proteins for later purification and separation or cleavage. Other suitable fusion proteins are those with β-galactosidase, ubiquitin, the like.

25 Promoters that are most commonly used in recombinant DNA construction include the β-lactamase (penicillinase), lactose and tryptophan (trp) promoter systems. While these are the most commonly used, other microbial promoters have been discovered and utilized, and details concerning their nucleotide sequences have been published, enabling those of skill in the art to ligate them functionally with plasmid vectors.

The following details concerning recombinant protein production in bacterial cells, such as *E. coli*, are provided by way of exemplary information on recombinant protein production in general, the adaptation of which to a particular recombinant expression system will be known to those of skill in the art.

5

Bacterial cells, for example, *E. coli*, containing the expression vector are grown in any of a number of suitable media, for example, LB. The expression of the recombinant protein may be induced, e.g., by adding IPTG to the media or by switching incubation to a higher temperature. After culturing the bacteria for a further period, generally of between 2 and 24 h, the cells are
10 collected by centrifugation and washed to remove residual media.

The bacterial cells are then lysed, for example, by disruption in a cell homogenizer and centrifuged to separate the dense inclusion bodies and cell membranes from the soluble cell components. This centrifugation can be performed under conditions whereby the dense inclusion bodies are selectively enriched by incorporation of sugars, such as sucrose, into the buffer and centrifugation at a selective speed.
15

If the recombinant protein is expressed in the inclusion bodies, as is the case in many instances, these can be washed in any of several solutions to remove some of the contaminating host proteins, then solubilized in solutions containing high concentrations of urea (e.g. 8M) or chaotropic agents such as guanidine hydrochloride in the presence of reducing agents, such as β -mercaptoethanol or DTT (dithiothreitol).
20

Under some circumstances, it may be advantageous to incubate the protein for several h
25 under conditions suitable for the protein to undergo a refolding process into a conformation which more closely resembles that of the native protein. Such conditions generally include low protein concentrations, less than 500 mg/ml, low levels of reducing agent, concentrations of urea less than 2 M and often the presence of reagents such as a mixture of reduced and oxidized glutathione which facilitate the interchange of disulfide bonds within the protein molecule.
30

The refolding process can be monitored, for example, by SDS-PAGE, or with antibodies specific for the native molecule (which can be obtained from animals vaccinated with the native molecule or smaller quantities of recombinant protein). Following refolding, the protein can then be purified further and separated from the refolding mixture by chromatography on any of 5 several supports including ion exchange resins, gel permeation resins or on a variety of affinity columns.

For expression in *Saccharomyces*, the plasmid YRp7, for example, is commonly used. This plasmid already contains the *trpl* gene which provides a selection marker for a mutant strain 10 of yeast lacking the ability to grow in tryptophan, for example ATCC No. 44076 or PEP4-1. The presence of the *trpl* lesion as a characteristic of the yeast host cell genome then provides an effective environment for detecting transformation by growth in the absence of tryptophan.

Suitable promoting sequences in yeast vectors include the promoters for 15 3-phosphoglycerate kinase or other glycolytic enzymes, such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase. In constructing suitable expression plasmids, the termination sequences associated with these genes are also ligated into the expression vector 3' 20 of the sequence desired to be expressed to provide polyadenylation of the mRNA and termination.

Other suitable promoters, which have the additional advantage of transcription controlled by growth conditions, include the promoter region for alcohol dehydrogenase 2, 25 isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and the aforementioned glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization.

In addition to microorganisms, cultures of cells derived from multicellular organisms may also be used as hosts. In principle, any such cell culture is workable, whether from vertebrate or invertebrate culture. In addition to mammalian cells, these include insect cell systems infected with recombinant virus expression vectors (e.g., baculovirus); and plant cell systems infected with recombinant virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (e.g., Ti plasmid) containing one or more P-TEFb or kinase protein coding sequences.

In a useful insect system, *Autograph californica* nuclear polyhidrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in *Spodoptera frugiperda* cells. The P-TEFb or kinase protein coding sequences are cloned into non-essential regions (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of the coding sequences results in the inactivation of the polyhedrin gene and production of non-occluded recombinant virus (*i.e.*, virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted gene is expressed (e.g., U.S. Patent No. 4,215,051, Smith, incorporated herein by reference).

Examples of useful mammalian host cell lines are VERO and HeLa cells, Chinese hamster ovary (CHO) cell lines, W138, BHK, COS-7, 293, HepG2, 3T3, RIN and MDCK cell lines. In addition, a host cell strain may be chosen that modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (e.g., glycosylation) and processing (e.g., cleavage) of protein products may be important for the function of the protein.

Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins. Appropriate cells lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells such as 293 cells have already been shown to produce active P-TEFb.

30

Expression vectors for use in mammalian such cells ordinarily include an origin of replication (as necessary), a promoter located in front of the gene to be expressed, along with any necessary ribosome binding sites, RNA splice sites, polyadenylation site, and transcriptional terminator sequences. The origin of replication may be provided either by construction of the vector to include an exogenous origin, such as may be derived from SV40 or other viral (e.g., Polyoma, Adeno, VSV, BPV) source, or may be provided by the host cell chromosomal replication mechanism. If the vector is integrated into the host cell chromosome, the latter is often sufficient.

The promoters may be derived from the genome of mammalian cells (e.g., metallothionein promoter) or from mammalian viruses (e.g., the adenovirus late promoter; the vaccinia virus 7.5K promoter). Further, it is also possible, and may be desirable, to utilize promoter or control sequences normally associated with the desired P-TEFb or kinase protein gene sequence, provided such control sequences are compatible with the host cell systems.

A number of viral based expression systems may be utilized, for example, commonly used promoters are derived from polyoma, Adenovirus 2, and most frequently Simian Virus 40 (SV40). The early and late promoters of SV40 virus are particularly useful because both are obtained easily from the virus as a fragment which also contains the SV40 viral origin of replication. Smaller or larger SV40 fragments may also be used, provided there is included the approximately 250 bp sequence extending from the *Hind III* site toward the *BgI I* site located in the viral origin of replication.

In cases where an adenovirus is used as an expression vector, the coding sequences may be ligated to an adenovirus transcription/ translation control complex, e.g., the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (e.g., region E1 or E3) will result in a recombinant virus that is viable and capable of expressing P-TEFb or kinase proteins in infected hosts.

Specific initiation signals may also be required for efficient translation of P-TEFb or kinase protein coding sequences. These signals include the ATG initiation codon and adjacent sequences. Exogenous translational control signals, including the ATG initiation codon, may additionally need to be provided. One of ordinary skill in the art would readily be capable of 5 determining this and providing the necessary signals. It is well known that the initiation codon must be in-frame (or in-phase) with the reading frame of the desired coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription 10 terminators.

In eukaryotic expression, one will also typically desire to incorporate into the transcriptional unit an appropriate polyadenylation site (*e.g.*, 5'-AATAAA-3') if one was not contained within the original cloned segment. Typically, the poly-A addition site is placed about 15 30 to 2000 nucleotides "downstream" of the termination site of the protein at a position prior to transcription termination.

For long-term, high-yield production of recombinant P-TEFb or kinase proteins, stable expression is preferred. For example, cell lines that stably express constructs encoding P-TEFb or kinase proteins may be engineered. Rather than using expression vectors that contain viral 20 origins of replication, host cells can be transformed with vectors controlled by appropriate expression control elements (*e.g.*, promoter, enhancer, sequences, transcription terminators, polyadenylation sites, *etc.*), and a selectable marker. Following the introduction of foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are 25 switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci which in turn can be cloned and expanded into cell lines.

A number of selection systems may be used, including, but not limited, to the herpes simplex virus (HSV) tk, hypoxanthine-guanine phosphoribosyltransferase (hgprt) and adenine phosphoribosyltransferase genes (aprt), in tk-, hgprt- or aprt- cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for *dhfr*, that confers resistance to methotrexate; *gpt*, that confers resistance to mycophenolic acid; *neo*, that confers resistance to the aminoglycoside G-418; and *hygro*, that confers resistance to hygromycin.

Animal cells can be propagated *in vitro* in two modes: as non-anchorage dependent cells growing in suspension throughout the bulk of the culture or as anchorage-dependent cells requiring attachment to a solid substrate for their propagation (*i.e.*, a monolayer type of cell growth).

Non-anchorage dependent or suspension cultures from continuous established cell lines are the most widely used means of large scale production of cells and cell products. However, suspension cultured cells have limitations, such as tumorigenic potential and lower protein production than adherent cells.

Large scale suspension culture of mammalian cells in stirred tanks is a common method for production of recombinant proteins. Two suspension culture reactor designs are in wide use - the stirred reactor and the airlift reactor. The stirred design has successfully been used on an 8000 liter capacity for the production of interferon. Cells are grown in a stainless steel tank with a height-to-diameter ratio of 1:1 to 3:1. The culture is usually mixed with one or more agitators, based on bladed disks or marine propeller patterns. Agitator systems offering less shear forces than blades have been described. Agitation may be driven either directly or indirectly by magnetically coupled drives. Indirect drives reduce the risk of microbial contamination through seals on stirrer shafts.

The airlift reactor, also initially described for microbial fermentation and later adapted for mammalian culture, relies on a gas stream to both mix and oxygenate the culture. The gas stream enters a riser section of the reactor and drives circulation. Gas disengages at the culture

surface, causing denser liquid free of gas bubbles to travel downward in the downcomer section of the reactor. The main advantage of this design is the simplicity and lack of need for mechanical mixing. Typically, the height-to-diameter ratio is 10:1. The airlift reactor scales up relatively easily, has good mass transfer of gases and generates relatively low shear forces.

5

It is contemplated that the P-TEFb subunits or proteins of the invention may be "overexpressed", *i.e.*, expressed in increased levels relative to its natural expression in cells. Such overexpression may be assessed by a variety of methods, including radio-labeling and/or protein purification. However, simple and direct methods are preferred, for example, those involving SDS/PAGE and protein staining or western blotting, followed by quantitative analyses, such as densitometric scanning of the resultant gel or blot. A specific increase in the level of the recombinant protein or peptide in comparison to the level in natural cells is indicative of overexpression, as is a relative abundance of the specific protein in relation to the other proteins produced by the host cell and, *e.g.*, visible on a gel.

15

III. P-TEFb Proteins and Peptides

The present invention therefore provides purified, and in preferred embodiments, substantially purified, P-TEFb proteins, subunits and peptides. The term "purified P-TEFb protein, subunit or peptide" as used herein, is intended to refer to a P-TEFb proteinaceous composition, isolatable from insect, mammalian, human or recombinant host cells, wherein the P-TEFb protein, subunit or peptide is prepared at levels that could not be previously obtained prior to the cloning of the P-TEFb subunit genes. A purified P-TEFb protein, subunit or peptide therefore also refers to a P-TEFb protein, subunit or peptide free from the environment in which it naturally occurs.

25

P-TEFb proteins may be full length proteins comprising either a small subunit or large subunit, such as being 404 (SEQ ID NO:2), 1113 (SEQ ID NO:4), 696 (SEQ ID NO:45), 729 (SEQ ID NO:47) or 726 (SEQ ID NO:50) amino acids in length. P-TEFb proteins, polypeptides and peptides may also be less than full length proteins, such as individual domains, regions or even epitopic peptides. Where less than full length P-TEFb proteins are concerned the most

preferred will be those containing predicted immunogenic sites and those containing the functional domains identified herein. Preferred P-TEFb kinase domains or fragments will be those sufficient to phosphorylate RNA polymerase II.

5 Generally, "purified" will refer to a P-TEFb protein, subunit or peptide composition that has been subjected to fractionation to remove various non-P-TEFb protein, subunit or peptide components, and which composition substantially retains its P-TEFb activity, as may be assessed by phosphorylation of RNA polymerase II and inhibition by DRB.

10 Where the term "substantially purified" is used, this will refer to a composition in which the P-TEFb protein, subunit or peptide forms the major component of the composition, such as constituting about 50% of the proteins in the composition or more. In preferred embodiments, a substantially purified protein will constitute more than 60%, 70%, 80%, 90%, 95%, 99% or even more of the proteins in the composition.

15 A polypeptide or protein that is "purified to homogeneity," as applied to the present invention, means that the polypeptide or protein has a level of purity where the polypeptide or protein is substantially free from other proteins and biological components. For example, a purified polypeptide or protein will often be sufficiently free of other protein components so that degradative sequencing may be performed successfully.

20 Various methods for quantifying the degree of purification of P-TEFb proteins, subunits or peptides will be known to those of skill in the art in light of the present disclosure. These include, for example, determining the specific ability to phosphorylate RNA polymerase II of a fraction, or assessing the number of polypeptides within a fraction by gel electrophoresis.

25 To purify a P-TEFb protein, subunit or peptide a natural or recombinant composition comprising at least some P-TEFb proteins, subunits or peptides will be subjected to fractionation to remove various non-P-TEFb components from the composition. Various techniques suitable for use in protein purification will be well known to those of skill in the art. These include, for

example, precipitation with ammonium sulfate, PEG, antibodies and the like or by heat denaturation, followed by centrifugation; chromatography steps such as ion exchange, gel filtration, reverse phase, hydroxylapatite, lectin affinity and other affinity chromatography steps; isoelectric focusing; gel electrophoresis; and combinations of such and other techniques.

5

A specific example presented herein is the purification of a P-TEFb fusion protein using a specific binding partner. Such purification methods are routine in the art. As the present invention provides DNA sequences for P-TEFb proteins, any fusion protein purification method can now be practiced. This is currently exemplified by the generation of a P-TEFb-glutathione
10 S-transferase fusion protein, expression in *E. coli*, and isolation to homogeneity using affinity chromatography on glutathione-agarose.

The exemplary purification method of fractionation *via* column chromatography, using a series of chromatography columns, such as phosphocellulose (P-11), DEAE cellulose, phenyl sepharose, hydroxylapatite, MonoQ, MonoS and the like, as disclosed herein represents one method to prepare a substantially purified P-TEFb protein, subunit or peptide. This method is preferred as it results in the substantial purification of the P-TEFb protein, subunit or peptide in yields sufficient for further characterization and use. However, given the DNA and proteins provided by the present invention, any purification method can now be employed.
15
20

Although preferred for use in certain embodiments, there is no general requirement that the P-TEFb protein, subunit or peptide always be provided in their most purified state. Indeed, it is contemplated that less substantially purified P-TEFb proteins, subunits or peptides, which are nonetheless enriched in P-TEFb compositions, relative to the natural state, will have utility in
25 certain embodiments. These include, for example, phosphorylating RNA polymerase II, as may be used to examine RNA elongation; and antibody generation where subsequent inhibitor screening assays using purified P-TEFb are conducted.

Methods exhibiting a lower degree of relative purification may have advantages in total recovery of protein product, or in maintaining the activity of an expressed protein. Inactive products also have utility in certain embodiments, such as, *e.g.*, in antibody generation.

5 **IV. P-TEFb Antibodies and Immunological Reagents**

A. **Epitopic Core Sequences**

Peptides corresponding to one or more antigenic determinants, or "epitopic core regions", of P-TEFb proteins or subunits of the present invention can also be prepared. Such peptides should generally be at least 7 or 8 amino acid residues in length, will preferably be about 10, 15, 10 20, 25 or about 30 amino acid residues in length, and may contain up to about 35-50 residues or so.

Synthetic peptides will generally be about, or less than about, 35-36 residues long, which is the approximate upper length limit of automated peptide synthesis machines, such as those available from Applied Biosystems (Foster City, CA). Longer peptides may also be prepared, *e.g.*, by recombinant means.

U.S. Patent 4,554,101, (Hopp) incorporated herein by reference, teaches the identification and preparation of epitopes from primary amino acid sequences on the basis of hydrophilicity. Through the methods disclosed in Hopp, one of skill in the art would be able to identify epitopes from within an amino acid sequence such as the P-TEFb sequence disclosed herein (SEQ ID NO:2; SEQ ID NO:4; SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50).

Numerous scientific publications have also been devoted to the prediction of secondary 25 structure, and to the identification of epitopes, from analyses of amino acid sequences (Chou and Fasman, 1974a,b; 1978a,b, 1979). Any of these may be used, if desired, to supplement the teachings of Hopp in U.S. Patent 4,554,101.

Moreover, computer programs are currently available to assist with predicting antigenic 30 portions and epitopic core regions of proteins. Examples include those programs based upon the

Jameson-Wolf analysis (Jameson and Wolf, 1998; Wolf *et al.*, 1988), the program PepPlot® (Brutlag *et al.*, 1990; Weinberger *et al.*, 1985), and other new programs for protein tertiary structure prediction (Fetrow and Bryant, 1993). Further commercially available software capable of carrying out such analyses is termed MacVector® (IBI, New Haven, CT).

5

In further embodiments, major antigenic determinants of a polypeptide may be identified by an empirical approach in which portions of the gene encoding the polypeptide are expressed in a recombinant host, and the resulting proteins tested for their ability to elicit an immune response. For example, PCR™ can be used to prepare a range of peptides lacking successively longer fragments of the C-terminus of the protein. The immunoactivity of each of these peptides is determined to identify those fragments or domains of the polypeptide that are immunodominant. Further studies in which only a small number of amino acids are removed at each iteration then allows the location of the antigenic determinants of the polypeptide to be more precisely determined.

15

Once one or more such analyses are completed, polypeptides are prepared that contain at least the essential features of one or more antigenic determinants. The peptides are then employed in the generation of antisera against the polypeptide. Minigenes or gene fusions encoding these determinants can also be constructed and inserted into expression vectors by standard methods, for example, using PCR™ cloning methodology.

20

The use of such small peptides for vaccination typically requires conjugation of the peptide to an immunogenic carrier protein, such as hepatitis B surface antigen, keyhole limpet hemocyanin or bovine serum albumin. Methods for performing this conjugation are well known in the art.

25

B. Antibody Generation

In certain embodiments, the present invention provides antibodies that bind with high specificity to P-TEFb, and other antibodies that bind to the protein products of the isolated

nucleic acid sequences of SEQ ID NO:1, SEQ ID NO:3 SEQ ID NO:43, SEQ ID NO:46 or SEQ ID NO:48. Antibodies specific for the wild type proteins and peptides and those specific for any one of a number of mutants are provided. As detailed above, in addition to antibodies generated against the full length proteins, antibodies may also be generated in response to smaller constructs comprising epitopic core regions, including wild type and mutant epitopes.

As used herein, the term "antibody" is intended to refer broadly to any immunologic binding agent such as IgG, IgM, IgA, IgD and IgE. Generally, IgG and/or IgM are preferred because they are the most common antibodies in the physiological situation and because they are most easily made in a laboratory setting.

Monoclonal antibodies (Mabs) are recognized to have certain advantages, e.g., reproducibility and large-scale production, and their use is generally preferred. The invention thus provides Mabs of the human, murine, monkey, rat, hamster, rabbit and even chicken origin. Due to the ease of preparation and ready availability of reagents, murine Mabs will often be preferred.

However, "humanized" antibodies are also contemplated, as are chimeric antibodies from mouse, rat, or other species, bearing human constant and/or variable region domains, bispecific antibodies, recombinant and engineered antibodies and fragments thereof. Methods for the development of antibodies that are "custom-tailored" to the patient's tumor are likewise known and such custom-tailored antibodies are also contemplated.

The term "antibody" is used to refer to any antibody-like molecule that has an antigen binding region, and includes antibody fragments such as Fab', Fab, F(ab')₂, single domain antibodies (DABs), Fv, scFv (single chain Fv), and the like. The techniques for preparing and using various antibody-based constructs and fragments are well known in the art. Means for preparing and characterizing antibodies are well known in the art (See, e.g., Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, 1988; incorporated herein by reference).

The methods for generating Mabs generally begin along the same lines as those for preparing polyclonal antibodies. Briefly, a polyclonal antibody is prepared by immunizing an animal with an immunogenic P-TEFb composition in accordance with the present invention and collecting antisera from that immunized animal.

5

A wide range of animal species can be used for the production of antisera. Typically the animal used for production of anti-antisera is a rabbit, a mouse, a rat, a hamster, a guinea pig or a goat. Because of the relatively large blood volume of rabbits, a rabbit is a preferred choice for production of polyclonal antibodies.

10

As is well known in the art, a given composition may vary in its immunogenicity. It is often necessary therefore to boost the host immune system, as may be achieved by coupling a peptide or polypeptide immunogen to a carrier. Exemplary and preferred carriers are keyhole limpet hemocyanin (KLH) and bovine serum albumin (BSA). Other albumins such as ovalbumin, mouse serum albumin or rabbit serum albumin can also be used as carriers. Means for conjugating a polypeptide to a carrier protein are well known in the art and include glutaraldehyde, m-maleimidobenzoyl-N-hydroxysuccinimide ester, carbodiimide and bis-biazotized benzidine.

PATENT
PENDING
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
22210
22211
22212
22213
22214
22215
22216
22217
22218
22219
22220
22221
22222
22223
22224
22225
22226
22227
22228
22229
222210
222211
222212
222213
222214
222215
222216
222217
222218
222219
222220
222221
222222
222223
222224
222225
222226
222227
222228
222229
2222210
2222211
2222212
2222213
2222214
2222215
2222216
2222217
2222218
2222219
2222220
2222221
2222222
2222223
2222224
2222225
2222226
2222227
2222228
2222229
22222210
22222211
22222212
22222213
22222214
22222215
22222216
22222217
22222218
22222219
22222220
22222221
22222222
22222223
22222224
22222225
22222226
22222227
22222228
22222229
222222210
222222211
222222212
222222213
222222214
222222215
222222216
222222217
222222218
222222219
222222220
222222221
222222222
222222223
222222224
222222225
222222226
222222227
222222228
222222229
2222222210
2222222211
2222222212
2222222213
2222222214
2222222215
2222222216
2222222217
2222222218
2222222219
2222222220
2222222221
2222222222
2222222223
2222222224
2222222225
2222222226
2222222227
2222222228
2222222229
22222222210
22222222211
22222222212
22222222213
22222222214
22222222215
22222222216
22222222217
22222222218
22222222219
22222222220
22222222221
22222222222
22222222223
22222222224
22222222225
22222222226
22222222227
22222222228
22222222229
222222222210
222222222211
222222222212
222222222213
222222222214
222222222215
222222222216
222222222217
222222222218
222222222219
222222222220
222222222221
222222222222
222222222223
222222222224
222222222225
222222222226
222222222227
222222222228
222222222229
2222222222210
2222222222211
2222222222212
2222222222213
2222222222214
2222222222215
2222222222216
2222222222217
2222222222218
2222222222219
2222222222220
2222222222221
2222222222222
2222222222223
2222222222224
2222222222225
2222222222226
2222222222227
2222222222228
2222222222229
22222222222210
22222222222211
22222222222212
22222222222213
22222222222214
22222222222215
22222222222216
22222222222217
22222222222218
22222222222219
22222222222220
22222222222221
22222222222222
22222222222223
22222222222224
22222222222225
22222222222226
22222222222227
22222222222228
22222222222229
222222222222210
222222222222211
222222222222212
222222222222213
222222222222214
222222222222215
222222222222216
222222222222217
222222222222218
222222222222219
222222222222220
222222222222221
222222222222222
222222222222223
222222222222224
222222222222225
222222222222226
222222222222227
222222222222228
222222222222229
2222222222222210
2222222222222211
2222222222222212
2222222222222213
2222222222222214
2222222222222215
2222222222222216
2222222222222217
2222222222222218
2222222222222219
2222222222222220
2222222222222221
2222222222222222
2222222222222223
2222222222222224
2222222222222225
2222222222222226
2222222222222227
2222222222222228
2222222222222229
22222222222222210
22222222222222211
22222222222222212
22222222222222213
22222222222222214
22222222222222215
22222222222222216
22222222222222217
22222222222222218
22222222222222219
22222222222222220
22222222222222221
22222222222222222
22222222222222223
22222222222222224
22222222222222225
22222222222222226
22222222222222227
22222222222222228
22222222222222229
222222222222222210
222222222222222211
222222222222222212
222222222222222213
222222222222222214
222222222222222215
222222222222222216
222222222222222217
222222222222222218
222222222222222219
222222222222222220
222222222222222221
222222222222222222
222222222222222223
222222222222222224
222222222222222225
222222222222222226
222222222222222227
222222222222222228
222222222222222229
2222222222222222210
2222222222222222211
2222222222222222212
2222222222222222213
2222222222222222214
2222222222222222215
2222222222222222216
2222222222222222217
2222222222222222218
2222222222222222219
2222222222222222220
2222222222222222221
2222222222222222222
2222222222222222223
2222222222222222224
2222222222222222225
2222222222222222226
2222222222222222227
2222222222222222228
2222222222222222229
22222222222222222210
22222222222222222211
22222222222222222212
22222222222222222213
22222222222222222214
22222222222222222215
22222222222222222216
22222222222222222217
22222222222222222218
22222222222222222219
22222222222222222220
22222222222222222221
22222222222222222222
22222222222222222223
22222222222222222224
22222222222222222225
22222222222222222226
22222222222222222227
22222222222222222228
22222222222222222229
222222222222222222210
222222222222222222211
222222222222222222212
222222222222222222213
222222222222222222214
222222222222222222215
222222222222222222216
222222222222222222217
222222222222222222218
222222222222222222219
222222222222222222220
222222222222222222221
222222222222222222222
222222222222222222223
222222222222222222224
222222222222222222225
222222222222222222226
222222222222222222227
222222222222222222228
222222222222222222229
2222222222222222222210
2222222222222222222211
2222222222222222222212
2222222222222222222213
2222222222222222222214
2222222222222222222215
2222222222222222222216
2222222222222222222217
2222222222222222222218
2222222222222222222219
2222222222222222222220
2222222222222222222221
2222222222222222222222
2222222222222222222223
2222222222222222222224
2222222222222222222225
2222222222222222222226
2222222222222222222227
2222222222222222222228
2222222222222222222229
22222222222222222222210
22222222222222222222211
22222222222222222222212
22222222222222222222213
22222222222222222222214
22222222222222222222215
22222222222222222222216
22222222222222222222217
22222222222222222222218
22222222222222222222219
22222222222222222222220
22222222222222222222221
22222222222222222222222
22222222222222222222223
22222222222222222222224
22222222222222222222225
22222222222222222222226
22222222222222222222227
22222222222222222222228
22222222222222222222229
222222222222222222222210
222222222222222222222211
222222222222222222222212
222222222222222222222213
222222222222222222222214
222222222222222222222215
222222222222222222222216
222222222222222222222217
222222222222222222222218
222222222222222222222219
222222222222222222222220
222222222222222222222221
222222222222222222222222
222222222222222222222223
222222222222222222222224
222222222222222222222225
222222222222222222222226
222222222222222222222227
222222222222222222222228
222222222222222222222229
2222222222222222222222210
2222222222222222222222211
2222222222222222222222212
2222222222222222222222213
2222222222222222222222214
2222222222222222222222215
2222222222222222222222216
2222222222222222222222217
2222222222222222222222218
2222222222222222222222219
2222222222222222222222220
2222222222222222222222221
2222222222222222222222222
2222222222222222222222223
2222222222222222222222224
2222222222222222222222225
2222222222222222222222226
2222222222222222222222227
2222222222222222222222228
2222222222222222222222229
22222222222222222222222210
22222222222222222222222211
22222222222222222222222212
22222222222222222222222213
22222222222222222222222214
22222222222222222222222215
22222222222222222222222216
22222222222222222222222217
22222222222222222222222218
22222222222222222222222219
22222222222222222222222220
22222222222222222222222221
22222222222222222222222222
22222222222222222222222223
22222222222222222222222224
22222222222222222222222225
22222222222222222222222226
22222222222222222222222227
22222222222222222222222228
22222222222222222222222229
222222222222222222222222210
222222222222222222222222211
222222222222222222222222212
222222222222222222222222213
222222222222222222222222214
222222222222222222222222215
222222222222222222222222216
222222222222222222222222217
222222222222222222222222218
222222222222222222222222219
222222222222222222222222220
222222222222222222222222221
222222222222222222222222222
222222222222222222222222223
222222222222222222222222224
222222222222222222222222225
222222222222222222222222226
222222222222222222222222227
222222222222222222222222228
222222222222222222222222229
2222222222222222222222222210
2222222222222222222222222211
2222222222222222222222222212
2222222222222222222222222213
2222222222222222222222222214
2222222222222222222222222215
2222222222222222222222222216
2222222222222222222222222217
2222222222222222222222222218
2222222222222222222222222219
2222222222222222222222222220
2222222222222222222222222221
2222222222222222222222222222
2222222222222222222222222223
2222222222222222222222222224
2222222222222222222222222225
2222222222222222222222222226
2222222222222222222222222227
2222222222222222222222222228
2222222222222222222222222229
22222222222222222222222222210
22222222222222222222222222211
22222222222222222222222222212
22222222222222222222222222213
22222222222222222222222222214
22222222222222222222222222215
22222222222222222222222222216
22222222222222222222222222217
22222222222222222222222222218
22222222222222222222222222219
22222222222222222222222222220
22222222222222222222222222221
22222222222222222222222222222
22222222222222222222222222223
22222222222222222222222222224
22222222222222222222222222225
22222222222222222222222222226
22222222222222222222222222227
22222222222222222222222222228
22222222222222222222222222229
222222222222222222222222222210
222222222222222222222222222211
22222222222222

containing killed *Mycobacterium tuberculosis*), incomplete Freund's adjuvants and aluminum hydroxide adjuvant.

In addition to adjuvants, it may be desirable to coadminister biologic response modifiers (BRM), which have been shown to upregulate T cell immunity or downregulate suppresser cell activity. Such BRMs include, but are not limited to, Cimetidine (CIM; 1200 mg/d) (Smith/Kline, PA); or low-dose Cyclophosphamide (CYP; 300 mg/m²) (Johnson/ Mead, NJ) and Cytokines such as g-interferon, IL-2, or IL-12 or genes encoding proteins involved in immune helper functions, such as B-7.

10

The amount of immunogen composition used in the production of polyclonal antibodies varies upon the nature of the immunogen as well as the animal used for immunization. A variety of routes can be used to administer the immunogen (subcutaneous, intramuscular, intradermal, intravenous and intraperitoneal). The production of polyclonal antibodies may be monitored by sampling blood of the immunized animal at various points following immunization.

15

A second, booster injection, may also be given. The process of boosting and titering is repeated until a suitable titer is achieved. When a desired level of immunogenicity is obtained, the immunized animal can be bled and the serum isolated and stored, and/or the animal can be used to generate MAbs.

20

25

For production of rabbit polyclonal antibodies, the animal can be bled through an ear vein or alternatively by cardiac puncture. The removed blood is allowed to coagulate and then centrifuged to separate serum components from whole cells and blood clots. The serum may be used as is for various applications or else the desired antibody fraction may be purified by well-known methods, such as affinity chromatography using another antibody, a peptide bound to a solid matrix, or by using, e.g., protein A or protein G chromatography.

30

MAbs may be readily prepared through use of well-known techniques, such as those exemplified in U.S. Patent 4,196,265, incorporated herein by reference. Typically, this technique

involves immunizing a suitable animal with a selected immunogen composition, *e.g.*, a purified or partially purified P-TEFb protein, polypeptide, peptide or domain, be it a wild type or mutant composition. The immunizing composition is administered in a manner effective to stimulate antibody producing cells.

5

The methods for generating MAbs generally begin along the same lines as those for preparing polyclonal antibodies. Rodents such as mice and rats are preferred animals, however, the use of rabbit, sheep frog cells is also possible. The use of rats may provide certain advantages (Goding, 1986, pp. 60-61), but mice are preferred, with the BALB/c mouse being
10 most preferred as this is most routinely used and generally gives a higher percentage of stable fusions.

The animals are injected with antigen, generally as described above. The antigen may be coupled to carrier molecules such as keyhole limpet hemocyanin if necessary. The antigen would typically be mixed with adjuvant, such as Freund's complete or incomplete adjuvant. Booster injections with the same antigen would occur at approximately two-week intervals.
15

Following immunization, somatic cells with the potential for producing antibodies, specifically B lymphocytes (B cells), are selected for use in the MAb generating protocol. These cells may be obtained from biopsied spleens, tonsils or lymph nodes, or from a peripheral blood sample. Spleen cells and peripheral blood cells are preferred, the former because they are a rich source of antibody-producing cells that are in the dividing plasmablast stage, and the latter because peripheral blood is easily accessible.
20

Often, a panel of animals will have been immunized and the spleen of animal with the highest antibody titer will be removed and the spleen lymphocytes obtained by homogenizing the spleen with a syringe. Typically, a spleen from an immunized mouse contains approximately 5×10^7 to 2×10^8 lymphocytes.
25

The antibody-producing B lymphocytes from the immunized animal are then fused with cells of an immortal myeloma cell, generally one of the same species as the animal that was immunized. Myeloma cell lines suited for use in hybridoma-producing fusion procedures preferably are non-antibody-producing, have high fusion efficiency, and enzyme deficiencies that render them incapable of growing in certain selective media which support the growth of only the desired fused cells (hybridomas).

Any one of a number of myeloma cells may be used, as are known to those of skill in the art (Goding, pp. 65-66, 1986; Campbell, pp. 75-83, 1984). For example, where the immunized animal is a mouse, one may use P3-X63/Ag8, X63-Ag8.653, NS1/1.Ag 4 1, Sp210-Ag14, FO, NSO/U, MPC-11, MPC11-X45-GTG 1.7 and S194/5XX0 Bul; for rats, one may use R210.RCY3, Y3-Ag 1.2.3, IR983F and 4B210; and U-266, GM1500-GRG2, LICR-LON-HMy2 and UC729-6 are all useful in connection with human cell fusions.

One preferred murine myeloma cell is the NS-1 myeloma cell line (also termed P3-NS-1-Ag4-1), which is readily available from the NIGMS Human Genetic Mutant Cell Repository by requesting cell line repository number GM3573. Another mouse myeloma cell line that may be used is the 8-azaguanine-resistant mouse murine SP2/0 non-producer cell line.

Methods for generating hybrids of antibody-producing spleen or lymph node cells and myeloma cells usually comprise mixing somatic cells with myeloma cells in a 2:1 proportion, though the proportion may vary from about 20:1 to about 1:1, respectively, in the presence of an agent or agents (chemical or electrical) that promote the fusion of cell membranes. Fusion methods using Sendai virus have been described by Kohler and Milstein (1975; 1976), and those using polyethylene glycol (PEG), such as 37% (v/v) PEG, by Gefter *et al.* (1977). The use of electrically induced fusion methods is also appropriate (Goding pp. 71-74, 1986).

Fusion procedures usually produce viable hybrids at low frequencies, about 1×10^{-6} to 1×10^{-8} . However, this does not pose a problem, as the viable, fused hybrids are differentiated

from the parental, unfused cells (particularly the unfused myeloma cells that would normally continue to divide indefinitely) by culturing in a selective medium. The selective medium is generally one that contains an agent that blocks the *de novo* synthesis of nucleotides in the tissue culture media. Exemplary and preferred agents are aminopterin, methotrexate, and azaserine.

5 Aminopterin and methotrexate block *de novo* synthesis of both purines and pyrimidines, whereas azaserine blocks only purine synthesis. Where aminopterin or methotrexate is used, the media is supplemented with hypoxanthine and thymidine as a source of nucleotides (HAT medium). Where azaserine is used, the media is supplemented with hypoxanthine.

10 The preferred selection medium is HAT. Only cells capable of operating nucleotide salvage pathways are able to survive in HAT medium. The myeloma cells are defective in key enzymes of the salvage pathway, *e.g.*, hypoxanthine phosphoribosyl transferase (HPRT), and they cannot survive. The B cells can operate this pathway, but they have a limited life span in culture and generally die within about two weeks. Therefore, the only cells that can survive in the selective media are those hybrids formed from myeloma and B cells.

15 This culturing provides a population of hybridomas from which specific hybridomas are selected. Typically, selection of hybridomas is performed by culturing the cells by single-clone dilution in microtiter plates, followed by testing the individual clonal supernatants (after about two to three weeks) for the desired reactivity. The assay should be sensitive, simple and rapid, such as radioimmunoassays, enzyme immunoassays, cytotoxicity assays, plaque assays, dot immunobinding assays, and the like.

20 The selected hybridomas would then be serially diluted and cloned into individual antibody-producing cell lines, which clones can then be propagated indefinitely to provide MAbs. The cell lines may be exploited for MAb production in two basic ways.

25 A sample of the hybridoma can be injected (often into the peritoneal cavity) into a histocompatible animal of the type that was used to provide the somatic and myeloma cells for the original fusion (*e.g.*, a syngeneic mouse). Optionally, the animals are primed with a

hydrocarbon, especially oils such as pristane (tetramethylpentadecane) prior to injection. The injected animal develops tumors secreting the specific monoclonal antibody produced by the fused cell hybrid. The body fluids of the animal, such as serum or ascites fluid, can then be tapped to provide MAbs in high concentration.

5

The individual cell lines could also be cultured *in vitro*, where the MAbs are naturally secreted into the culture medium from which they can be readily obtained in high concentrations.

MAbs produced by either means may be further purified, if desired, using filtration, 10 centrifugation and various chromatographic methods such as HPLC or affinity chromatography. Fragments of the MAbs of the invention can be obtained from the MAbs so produced by methods which include digestion with enzymes, such as pepsin or papain, and/or by cleavage of disulfide bonds by chemical reduction. Alternatively, monoclonal antibody fragments encompassed by the present invention can be synthesized using an automated peptide synthesizer.

15 It is also contemplated that a molecular cloning approach may be used to generate monoclonals. For this, combinatorial immunoglobulin phagemid libraries are prepared from RNA isolated from the spleen of the immunized animal, and phagemids expressing appropriate antibodies are selected by panning using cells expressing the antigen and control cells. The advantages of this approach over conventional hybridoma techniques are that approximately 10^4 times as many antibodies can be produced and screened in a single round, and that new 20 specificities are generated by H and L chain combination which further increases the chance of finding appropriate antibodies.

25 Alternatively, monoclonal antibody fragments encompassed by the present invention can be synthesized using an automated peptide synthesizer, or by expression of full-length gene or of gene fragments in *E. coli*.

C. Antibody Conjugates

The present invention further provides anti-P-TEFb antibodies, generally of the monoclonal type, that are linked to one or more other agents to form an antibody conjugate. Any antibody of sufficient selectivity, specificity and affinity may be employed as the basis for an antibody conjugate. Such properties may be evaluated using conventional immunological screening methodology known to those of skill in the art.

Certain examples of antibody conjugates are those conjugates in which the antibody is linked to a detectable label. "Detectable labels" are compounds or elements that can be detected due to their specific functional properties, or chemical characteristics, the use of which allows the antibody to which they are attached to be detected, and further quantified if desired. Another such example is the formation of a conjugate comprising an antibody linked to a cytotoxic or anti-cellular agent, as may be termed "immunotoxins". In the context of the present invention, immunotoxins are generally less preferred.

Antibody conjugates are thus preferred for use as diagnostic agents. Antibody diagnostics generally fall within two classes, those for use in *in vitro* diagnostics, such as in a variety of immunoassays, and those for use *in vivo* diagnostic protocols, generally known as "antibody-directed imaging". Again, antibody-directed imaging is less preferred for use with this invention.

Many appropriate imaging agents are known in the art, as are methods for their attachment to antibodies (see, e.g., U.S. patents 5,021,236 and 4,472,509, both incorporated herein by reference). Certain attachment methods involve the use of a metal chelate complex employing, for example, an organic chelating agent such a DTPA attached to the antibody (U.S. Patent 4,472,509). Monoclonal antibodies may also be reacted with an enzyme in the presence of a coupling agent such as glutaraldehyde or periodate. Conjugates with fluorescein markers are prepared in the presence of these coupling agents or by reaction with an isothiocyanate.

In the case of paramagnetic ions, one might mention by way of example ions such as chromium (III), manganese (II), iron (III), iron (II), cobalt (II), nickel (II), copper (II), neodymium (III), samarium (III), ytterbium (III), gadolinium (III), vanadium (II), terbium (III), dysprosium (III), holmium (III) and erbium (III), with gadolinium being particularly preferred.

5 Ions useful in other contexts, such as X-ray imaging, include but are not limited to lanthanum (III), gold (III), lead (II), and especially bismuth (III).

In the case of radioactive isotopes for therapeutic and/or diagnostic application, one might mention astatine²¹¹, ¹⁴carbon, ⁵¹chromium, ³⁶chlorine, ⁵⁷cobalt, ⁵⁸cobalt, copper⁶⁷, ¹⁵²Eu, gallium⁶⁷, ³hydrogen, iodine¹²³, iodine¹²⁵, iodine¹³¹, indium¹¹¹, ⁵⁹iron, ³²phosphorus, rhenium¹⁸⁶, rhenium¹⁸⁸, ⁷⁵selenium, ³⁵sulphur, technetium^{99m} and yttrium⁹⁰. ¹²⁵Iodine is often being preferred for use in certain embodiments, and technetium^{99m} and indium¹¹¹ are also often preferred due to their low energy and suitability for long range detection.

15 Radioactively labeled Mabs of the present invention may be produced according to well-known methods in the art. For instance, Mabs can be iodinated by contact with sodium or potassium iodide and a chemical oxidizing agent such as sodium hypochlorite, or an enzymatic oxidizing agent, such as lactoperoxidase. Monoclonal antibodies according to the invention may be labeled with technetium^{99m} by ligand exchange process, for example, by reducing pertechnate with stannous solution, chelating the reduced technetium onto a Sephadex column and applying the antibody to this column or by direct labeling techniques, e.g., by incubating pertechnate, a reducing agent such as SNCl₂, a buffer solution such as sodium-potassium phthalate solution, and the antibody.

20 25 Intermediary functional groups which are often used to bind radioisotopes which exist as metallic ions to antibody are diethylenetriaminepentaacetic acid (DTPA) and ethylene diaminetetracetic acid (EDTA). Fluorescent labels include rhodamine, fluorescein isothiocyanate and renographin.

The much preferred antibody conjugates of the present invention are those intended primarily for use *in vitro*, where the antibody is linked to a secondary binding ligand or to an enzyme (an enzyme tag) that will generate a colored product upon contact with a chromogenic substrate. Examples of suitable enzymes include urease, alkaline phosphatase, (horseradish) 5 hydrogen peroxidase and glucose oxidase. Preferred secondary binding ligands are biotin and avidin or streptavidin compounds. The use of such labels is well known to those of skill in the art in light and is described, for example, in U.S. Patents 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241; each incorporated herein by reference.

10

D. Immunodetection Methods and Kits

15

In still further embodiments, the present invention concerns immunodetection methods for binding, purifying, removing, quantifying or otherwise generally detecting biological components, such as HIV Tat, that interact with P-TEFb. The binding assays will often be used as a pre-screen to identify agents that inhibit productive RNA elongation, and especially to identify agents that inhibit viral-mediated RNA elongation by disrupting the interaction of viral transactivating proteins, such as Tat, that need to interact with P-TEFb in order to exert their effect on the host cell RNA polymerase II. Subsequent to the identification of components that inhibit the binding of a viral protein to P-TEFb in an immunodetection assay, the compounds will generally be further examined in RNA elongation activity assays to confirm their suitability.

20

Immunodetection methods may also be employed to identify novel proteins or protein domains that bind to P-TEFb, and/or to identify a new property of a known protein, allowing the definition of that protein as a P-TEFb-binding protein. In any event, the identification of an additional viral protein that binds to P-TEFb will allow the identified protein to become the target of new anti-viral strategies. The first step of such anti-viral approaches will, again, often be based upon a binding assay in which potential anti-viral compounds are tested for their ability 25 to inhibit the binding of the newly discovered protein to P-TEFb.

25

Accordingly, in the following description of binding assays, the term "viral 30 transactivating protein" is used to refer to any known or discovered viral transactivating protein

or domain that binds to P-TEFb and, most preferably, that binds to human P-TEFb. Examples of viral transactivating proteins that are herein discovered to bind to P-TEFb are HIV Tat and VP16 from herpes virus. E1A from adenovirus is also contemplated to be an appropriate target for use in such a binding assay. Naturally, where the intent of the assay is to identify viral proteins that
5 bind to P-TEFb, the "viral transactivating protein" will more properly be termed a "viral composition comprising a candidate transactivating protein", which is generally exemplified by extracts comprising viral proteins produced during host cell infection.

Antibodies and labeled antibodies against the P-TEFb proteins, subunits or peptides of
10 the present invention may also be employed to purify P-TEFb, as may be used in purification from insect, mammalian or human sources, including recombinant host cells, *via* immunoaffinity protocols, such as immunoaffinity chromatography. In the purification methods, the antibody removes or purifies the antigenic P-TEFb component from a sample. The antibody will preferably be linked to a solid support, such as in the form of a column matrix, and the sample suspected of containing the P-TEFb antigenic component will be applied to the immobilized antibody. The unwanted components will be washed from the column, leaving the antigen immunocomplexed to the immobilized antibody, which P-TEFb antigen is then collected by
15 removing the P-TEFb from the column.

Antibodies and labeled antibodies against the P-TEFb proteins, subunits or peptides of
20 the invention may even further be employed to detect P-TEFb in samples, including recombinant host cells and clinical samples. Such antibodies may ultimately be employed in diagnostic embodiments to detect increased or decreased levels of P-TEFb or to detect mutant P-TEFb proteins, subunits or peptides in human tissue or fluid samples. As P-TEFb is centrally
25 important to RNA elongation, disturbances in P-TEFb levels or types are likely to be implicated in diseases, such as cancer, in which gene expression is perturbed. Therefore, the use of wild type- and mutant-specific antibodies is contemplated. However, these embodiments also have applications to non-clinical samples, such as in the titering of antigen or antibody samples, in the selection of hybridomas, and the like.

30

Accordingly, detection of P-TEFb alone is another important aspect of the invention. In general, simple P-TEFb immunobinding methods include obtaining a sample suspected of containing a P-TEFb protein, subunit, peptide, or mutant thereof, and contacting the sample with a first anti-P-TEFb antibody or anti-mutant P-TEFb antibody in accordance with the present invention, under conditions effective to allow the formation of immunocomplexes, and then detecting the immunocomplexes so formed.

In the clinical diagnosis or monitoring of patients with diseases such as cancer, the detection of a P-TEFb mutant, or an alteration in the levels of P-TEFb, in comparison to the levels in a corresponding biological sample from a normal subject will be indicative of a patient with the disease or cancer. However, as is known to those of skill in the art, such a clinical diagnosis would not necessarily be made on the basis of this method in isolation. Those of skill in the art are very familiar with differentiating between significant differences in types or amounts of biomarkers, which represent a positive identification, and low level or background changes of biomarkers. Indeed, background expression levels are often used to form a "cut-off" above which increased detection will be scored as significant or positive.

The P-TEFb-viral protein immunobinding methods of the present invention generally include obtaining a sample suspected of containing a viral transactivating protein, domain, subunit or active peptide, and contacting the sample with a first P-TEFb composition, complex, protein, subunit, domain or binding peptide, in accordance with the present invention, under conditions effective to allow the formation of bound protein complexes, and then detecting the bound protein complexes so formed. The detection of the "bound protein complexes", in the manner described herein, is highly analogous to the techniques of immunocomplex detection, even in embodiments where either the viral transactivating protein or the P-TEFb protein are directly labeled, and all such technology regarding immobilization, specific binding, non-specific binding and washing is applicable.

The steps of various useful immunodetection methods have been described in the scientific literature, such as, e.g., Nakamura *et al.* (1987), incorporated herein by reference. In

terms of detecting or quantifying the amount of viral transactivating proteins, *i.e.*, P-TEFb reactive components in a sample, the methods require the detection or quantification of any bound protein or immune complexes formed during the binding process. Here, one would obtain a sample known or suspected to contain a viral transactivating protein and contact the sample
5 with a P-TEFb protein, subunit or peptide, and then detect or quantify the amount of bound protein complexes formed under the specific conditions.

Contacting the chosen viral transactivating sample with the P-TEFb sample will be effected under conditions effective and for a period of time sufficient to allow the formation of
10 bound protein complexes (primary complexes), which is generally a matter of simply adding the two compositions or samples and incubating the mixture for a period of time long enough for any viral transactivating proteins and P-TEFb proteins present to form bound protein complexes. After this time, the potentially bound protein complexes, as may be present on an ELISA plate, dot blot or such like, will generally be washed to remove any non-specifically bound protein species, allowing only those species specifically bound within the primary complexes to be
15 detected.

In general, the detection of bound protein complexes is analogous to the detection of immunocomplexes, and is well known to those of skill in the art. Detection may be achieved through the application of numerous approaches. These methods are generally based upon the detection of a label or marker, such as any of those radioactive, fluorescent, biological or enzymatic tags. U.S. Patents concerning the use of such labels include 3,817,837; 3,850,752;
20 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241, each incorporated herein by reference.
25

Preferably, in detection, one would find additional advantages through the use of an antibody, or other secondary binding ligand, such as a second biotin/avidin ligand binding arrangement, as is known in the art. The antibody employed in the detection may be an anti-P-TEFb antibody or an anti-viral protein antibody, such as an anti-HIV Tat antibody. Such
30 antibodies would be termed "primary antibodies". The primary antibodies may themselves be

linked to a detectable label, wherein one would then simply detect this label, thereby allowing the amount of the bound protein complexes in the composition to be determined.

Alternatively, the first anti-P-TEFb antibody or anti-viral protein antibody (such as anti-HIV Tat) that becomes bound within the protein complexes may be detected by means of a second binding ligand that has binding affinity for the first antibody. In these cases, the second binding ligand may be linked to a detectable label. The second binding ligand is itself often an antibody, which may thus be termed a "secondary" antibody. The primary bound protein-antibody complexes are contacted with the labeled, secondary binding ligand, or antibody, under conditions effective and for a period of time sufficient to allow the formation of secondary immune complexes. The secondary immune complexes are then generally washed to remove any non-specifically bound labeled secondary antibodies or ligands, and the remaining label in the secondary immune complexes is then detected. Further methods including the use of tertiary binding ligands or antibodies linked to a detectable labels are also contemplated, particularly where signal amplification is desired.

1. ELISAs

As detailed above, immunoassays, in their most simple and direct sense, are binding assays. Certain preferred immunoassays are the various types of enzyme linked immunosorbent assays (ELISAs) and radioimmunoassays (RIA) known in the art. Immunohistochemical detection using tissue sections is also particularly useful. However, it will be readily appreciated that detection is not limited to such techniques, and western blotting, dot blotting, FACS analyses, and the like may also be used.

In one exemplary ELISA, either P-TEFb or a viral transactivating protein, such as HIV Tat, is immobilized onto a selected surface exhibiting protein affinity, such as a well in a polystyrene microtiter plate. Then, a composition containing the counterpart viral transactivating protein, or P-TEFb, is added to the wells. After binding and washing to remove non-specifically bound complexes, the bound P-TEFb-viral protein complex may be detected. Detection is generally achieved by the addition of an anti-P-TEFb or anti-viral protein antibody that is linked

to a detectable label. Detection may also be achieved by the addition of a first anti-P-TEFb or anti-viral protein antibody, followed by a second antibody that has binding affinity for the first antibody, with the second antibody being linked to a detectable label.

5 Irrespective of the format employed, ELISAs have certain features in common, such as coating, incubating or binding, washing to remove non-specifically bound species, and detecting the bound immune complexes. These are described as follows:

10 In coating a plate with either P-TEFb or a viral transactivating protein, such as Tat, VP16 or E1A, one will generally incubate the wells of the plate with a solution of the agent, either overnight or for a specified period of hours. The wells of the plate will then be washed to remove incompletely adsorbed material. Any remaining available surfaces of the wells are then "coated" with a nonspecific protein that is neutral with regard to binding to the biological components. These include bovine serum albumin (BSA), casein and solutions of milk powder. 15 The coating allows for blocking of nonspecific adsorption sites on the immobilizing surface and thus reduces the background caused by nonspecific binding of proteins onto the surface.

20 In the ELISAs of the present invention it will probably be more customary to use a secondary or tertiary detection means rather than a direct procedure using a labeled P-TEFb or a viral transactivating protein. Thus, after binding of the first protein to the well, coating with a non-reactive material to reduce background, and washing to remove unbound material, the immobilizing surface is contacted with the second biological protein under conditions effective to allow protein complex formation. Detection of the complex then requires a labeled binding ligand or antibody.

25 "Under conditions effective to allow protein complex formation" means that the conditions preferably include diluting the P-TEFb and viral transactivating proteins, with solutions such as BSA, bovine gamma globulin (BGG) and phosphate buffered saline (PBS)/Tween. These added agents also tend to assist in the reduction of nonspecific background. 30 The "suitable" conditions also mean that the incubation is at a temperature and for a period of

time sufficient to allow effective binding. Incubation steps are typically from about 1 to 2 to 4 hours, at temperatures preferably on the order of 25° to 27°C, or may be overnight at about 4°C or so.

5 Following all incubation steps in an ELISA, the contacted surface is washed so as to remove non-complexed material. A preferred washing procedure includes washing with a solution such as PBS/Tween, or borate buffer. Following the formation of specific complexes between the test sample and the originally bound material, and subsequent washing, the occurrence of even minute amounts of bound complexes may be determined.

10

To provide for detection, a first or second antibody will preferably be provided that has an associated label to allow detection. However, given that the present invention provides highly purified P-TEFb components, and as viral transactivating proteins, such as HIV Tat, are readily available, the P-TEFb or viral protein may also be directly labeled.

15

Preferably, the label will be an enzyme that will generate color development upon incubating with an appropriate chromogenic substrate. Thus, for example, one will desire to contact and incubate the bound complexes with a urease, glucose oxidase, alkaline phosphatase or hydrogen peroxidase-conjugated antibody for a period of time and under conditions that favor the development of immunocomplex formation (*e.g.*, incubation for 2 hours at room temperature in a PBS-containing solution such as PBS-Tween).

20
25 After incubation with the labeled antibody, and subsequent to washing to remove unbound material, the amount of label is quantified, *e.g.*, by incubation with a chromogenic substrate such as urea and bromocresol purple or 2,2'-azino-di-(3-ethyl-benzthiazoline-6-sulfonic acid [ABTS] and H₂O₂, in the case of peroxidase as the enzyme label. Quantification is then achieved by measuring the degree of color generation, *e.g.*, using a visible spectra spectrophotometer.

In still further embodiments, the present invention concerns immunodetection kits for use with the immunodetection methods described above. P-TEFb, a viral transactivating protein and, preferably, a labeled antibody to at least one of such components will most generally be included in the kit. However, kits including less than or more than the foregoing components may also be provided. In preferred embodiments of the methods and kits, the antibodies used will be monoclonal antibodies (MAbs).

The immunodetection kits will comprise all the supplied components in suitable container means. In certain embodiments, the kits may comprise a P-TEFb or a viral transactivating protein, that is pre-bound to a solid support, such as a column matrix or well of a microtitre plate.

The immunodetection reagents of the kit may take any one of a variety of forms, including those detectable labels that are associated with or linked to the given P-TEFb or viral transactivating protein. Detectable labels that are associated with or attached to a secondary binding ligands or antibody are generally preferred. As noted above, a number of exemplary labels are known in the art and all such labels may be employed in connection with the present invention.

The kits may further comprise suitably aliquoted compositions of a P-TEFb protein or polypeptide or a viral transactivating protein, whether labeled or unlabeled, as may be used to prepare a standard curve for a detection assay. The kits may also contain protein- or antibody-label conjugates either in fully conjugated form, in the form of intermediates, or as separate moieties to be conjugated by the user of the kit. The components of the kits may be packaged either in aqueous media or in lyophilized form.

The container means of the kits will generally include at least one vial, test tube, flask, bottle, syringe or other container means, into which the P-TEFb or viral transactivating protein, and other optional components, may be placed, and preferably, suitably aliquoted. Where second and third additional binding components are provided, the kit will also generally contain a

second, third or other additional container into which such components may be placed. The kits of the present invention will also typically include a means for containing the reagent containers in close confinement for commercial sale. Such containers may include injection or blow-molded plastic containers into which the desired vials are retained.

5

V. Biological Functional Equivalents

As will be understood by those of skill in the art, modification and changes may be made in the structure of the P-TEFb kinase and large subunits and still obtain a molecule having like or otherwise desirable characteristics. For example, certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on molecules such as Tat and RNA polymerase II. Since it is the interactive capacity and nature of a protein that defines that protein's biological functional activity, certain amino acid sequence substitutions can be made in a protein sequence (or, of course, its underlying DNA coding sequence) and nevertheless obtain a protein with like (agonistic) properties. It is thus contemplated by the inventor that various changes may be made in the sequence of P-TEFb proteins or peptides (or underlying DNA) without appreciable loss of their biological utility or activity.

Equally, the same considerations may be employed to create a P-TEFb protein or peptide with countervailing (*e.g.*, antagonistic) properties. This is relevant to the present invention in which P-TEFb analogues without Tat binding activity are contemplated to be useful in inhibiting the ability of the HIV virus to promote viral RNA elongation by way of inhibiting Tat binding to P-TEFb.

In terms of functional equivalents, it is also well understood by the skilled artisan that, inherent in the definition of a biologically functional equivalent protein or peptide, is the concept that there is a limit to the number of changes that may be made within a defined portion of the molecule and still result in a molecule with an acceptable level of equivalent biological activity. Biologically functional equivalent peptides are thus defined herein as those peptides in which

certain, not most or all, of the amino acids may be substituted. In particular, where small peptides are concerned, less amino acids may be changed. Of course, a plurality of distinct proteins/peptides with different substitutions may easily be made and used in accordance with the invention.

5

It is also well understood that where certain residues are shown to be particularly important to the biological or structural properties of a protein or peptide, e.g., residues in the active site of an enzyme, or in the RNA polymerase II binding region, such residues may not generally be exchanged. This is the case in the present invention, where residues in the active 10 site of the kinase subunit should not generally be changed where it is the intention to maintain kinase function.

Amino acid substitutions are generally based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. An analysis of the size, shape and type of the amino acid side-chain substituents reveals that arginine, lysine and histidine are all positively charged residues; that alanine, glycine and serine are all a similar size; and that phenylalanine, tryptophan and tyrosine all have a generally similar shape. Therefore, based upon these considerations, arginine, lysine and histidine; alanine, glycine and serine; and phenylalanine, tryptophan and tyrosine; are defined herein as biologically functional equivalents.

To effect more quantitative changes, the hydropathic index of amino acids may be considered. Each amino acid has been assigned a hydropathic index on the basis of their hydrophobicity and charge characteristics, these are: isoleucine (+4.5); valine (+4.2); leucine 25 (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and arginine (-4.5).

The importance of the hydropathic amino acid index in conferring interactive biological function on a protein is generally understood in the art (Kyte and Doolittle, 1982, incorporated herein by reference). It is known that certain amino acids may be substituted for other amino acids having a similar hydropathic index or score and still retain a similar biological activity. In 5 making changes based upon the hydropathic index, the substitution of amino acids whose hydropathic indices are within ± 2 is preferred, those which are within ± 1 are particularly preferred, and those within ± 0.5 are even more particularly preferred.

It is also understood in the art that the substitution of like amino acids can be made 10 effectively on the basis of hydrophilicity, particularly where the biological functional equivalent protein or peptide thereby created is intended for use in immunological embodiments, as in the present case. U.S. Patent 4,554,101, incorporated herein by reference, states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with its immunogenicity and antigenicity, *i.e.* with a biological property of the 15 protein.

As detailed in U.S. Patent 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate ($+3.0 \pm 1$); glutamate ($+3.0 \pm 1$); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); 20 proline (-0.5 ± 1); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4).

In making changes based upon similar hydrophilicity values, the substitution of amino acids whose hydrophilicity values are within ± 2 is preferred, those which are within ± 1 are 25 particularly preferred, and those within ± 0.5 are even more particularly preferred.

While discussion has focused on functionally equivalent polypeptides arising from amino acid changes, it will be appreciated that these changes may be effected by alteration of the encoding DNA; taking into consideration also that the genetic code is degenerate and that two or 30 more codons may code for the same amino acid. A table of amino acids and their codons is

presented hereinabove for use in such embodiments, as well as for other uses, such as in the design of probes and primers and the like.

VI. Inhibitors and Screening Assays

5 In still further embodiments, the present invention provides methods for identifying new P-TEFb inhibitory compounds, which may be termed as "candidate substances." It is contemplated that such screening techniques will prove useful in the general identification of any compound that will serve the purpose of inhibiting P-TEFb, and in preferred embodiments, will provide candidate anti-viral compounds.

10

It is further contemplated that useful compounds in this regard will in no way be limited to proteinaceous or peptidyl compounds. In fact, it may prove to be the case that the most useful pharmacological compounds for identification through application of the screening assays will be non-peptidyl in nature and, *e.g.*, which will serve to inhibit viral RNA elongation through a tight binding or other chemical interaction. Candidate substances may be obtained from libraries of synthetic chemicals, or from natural samples, such as rain forest and marine samples.

A. Inhibition of P-TEFb Phosphorylation

RNA polymerase II is the natural substrate of the P-TEFb kinase and enzyme complex. 20 Although phosphorylation of the CTD of RNA polymerase II is complicated in cellular terms, using phosphorylation of RNA polymerase II as an assay to identify inhibitors of P-TEFb will be straightforward in light of the methods disclosed herein. The inhibitors initially identified in such assays can be used as general transcription elongation inhibitors, or may be modified to prepare second generation compounds for use as specific viral transcription elongation inhibitors.

25

General transcription elongation inhibitors will have utility in cellular assays, and are also contemplated for therapeutic uses. For example, the ability to generally inhibit transcription elongation is envisioned to be useful in controlling excessive gene transcription/translation and cellular proliferation, as may be used to treat cancer and other diseases associated with aberrant 30 gene expression and/or cellular reproduction and proliferation. It is well understood in the art

that many anti-cancer therapeutics, as well as anti-viral therapeutics, can exert certain effects in normal cell types, but such potential side-effects do not generally limit the therapeutic utility of such drugs. Further, the impact of any possible adverse effects can be limited or otherwise controlled by the more specific administration of the inhibitory agent to a disease localized site
5 or area of the body, such as by direct application to a tumor.

To identify a P-TEFb kinase inhibitor and potential transcription elongation inhibitor using a P-TEFb RNA polymerase II phosphorylation assay, one would simply conduct parallel or otherwise comparatively controlled phosphorylation assays and identify a compound that inhibits phosphorylation. The candidate screening assay is quite simple to set up and perform. Thus,
10 after obtaining a relatively purified preparation of the kinase or intact P-TEFb enzyme, either from native or recombinant sources, and either from *Drosophila* or human sources, one will simply admix a candidate substance with the kinase preparation, under conditions that would allow the enzyme to perform its P-TEFb function but for inclusion of an inhibitory substance.
15

For example, one will typically desire to include within the admixture an amount of a RNA polymerase II, although other phosphate acceptors may be used, such as phosphorlable peptides. Potential effectors, such as an HIV Tat protein, may also be included in the assay. In any event, one would measure the ability of the candidate substance to reduce phosphorylation of the RNA polymerase II or phosphate acceptor substrate relatively in the presence of the candidate substance. In general, one will desire to measure or otherwise determine the activity of the relatively purified enzyme in the absence of the added candidate substance relative to the activity in the presence of the candidate substance in order to assess the relative inhibitory capability of the candidate substance.
20
25

In terms of optimizing phosphorylation-based inhibitor assays, one should consider the following scientific observations. Each phosphorylation site is unique and may be influenced by phosphorylation of other sites. It is possible to saturate the assay with kinase such that even in the presence of inhibitors, all large subunit molecules are completely shifted to the I₁₀ form. In
30 addition, the exact relationship of the mobility of the large subunit to the number of phosphates

added is not known. The levels of phosphorylation required for function during transcription should be determined and the effect that other factors might have on the kinase action in the elongation complex should also be considered.

5 Therefore, the best correlation between kinase activity and function during transcription can be made by comparing the effects of two inhibitors under the same set of conditions. This does not pose a problem for conducting the P-TEFb phosphorylation-based inhibitor assays of the invention as DRB, and even H-8, are provided as known kinase inhibitors that also inhibit productive transcription elongation. Accordingly, in performing a P-TEFb phosphorylation-
10 based inhibitor assay, one may advantageously compare the effect of a candidate inhibitor with the effect of DRB, or even H-8, under the same set of conditions.

Prior to the present invention, there was continued confusion in the literature concerning the target for the action of the kinase inhibitor, DRB. DRB inhibits the CTD kinase activity of *Drosophila* P-TEFb and its ability to phosphorylate the CTD in EECs. The inventor has shown that under identical conditions the CTD kinase activity of P-TEFb is inhibited by DRB while that of TFIIH is unaffected by the drug. Yankulov *et. al.* (1996) proposed that the effect of DRB on transcription was due to inhibition of the kinase activity of TFIIH (Yankulov *et al.*, 1995). The Yankulov conclusions were based on the observations that DRB inhibited the TFIIH kinase and the appearance of runoff transcripts during transcription with similar dose-response curves. The Yankulov kinase and transcription assays were performed under conditions that did not allow a valid comparison of inhibition curves to be made. Although recent studies strongly implicate TFIIH in elongation control (Yankulov *et al.*, 1996), results with the *Drosophila* system indicate that TFIIH cannot substitute for the role of P-TEFb in controlling elongation.
20

25 The DRB-sensitive process whereby RNA polymerase II escapes abortive elongation and enters a productive mode of elongation is important in controlling transcription. The drugs DRB or H-8 inhibit the transition into productive elongation as well as the CTD kinase activity of P-TEFb. In addition, the present invention discloses that pure P-TEFb can phosphorylate the
30 CTD of RNA polymerase II in early elongation complexes. Post-initiation phosphorylation of

the CTD by P-TEFb controls the transition into productive elongation. P-TEFb may remain associated with the complex or it may be released, but the complex is DRB resistant after that point (Kephart *et al.*, 1992).

5 The present invention further discloses that the concentration of ATP used in phosphorylation assays has a direct effect on the concentration of inhibitor required to achieve 50% inhibition. This should be considered in conducting phosphorylation assays, and in correlating kinase assays, which generally use low ATP concentrations with transcription assays, which have higher levels of all NTPs. Those of ordinary skill in the art will understand that the
10 concentrations of the polymerase and kinase, the time of the reactions and other assay conditions, should be carefully monitored in analyzing the influence of potential inhibitors.

B. Inhibition of P-TEFb-Viral Protein Binding

Potential viral transcription elongation inhibitors can also be identified by conducting a
15 P-TEFb-viral protein binding assay, such as any of the assays described herein. To identify a potential viral transcription elongation inhibitor using a P-TEFb-viral protein binding assay also simply only requires one to conduct parallel or otherwise comparatively controlled binding assays and to identify a compound that inhibits P-TEFb-viral protein binding.

20 Such candidate inhibitor screening assays are also simple to set up and perform. One will simply admix a candidate substance with the P-TEFb and viral protein combination, under conditions that would normally allow the P-TEFb and viral proteins to bind but for inclusion of an inhibitory substance. One measures the ability of the candidate substance to reduce binding of the P-TEFb and viral proteins relatively in the presence of the candidate substance. In general,
25 one again measures or otherwise determines, *e.g.*, from a known standard curve, the binding of the same amount of the proteins in the absence of the added candidate substance relative to the binding in the presence of the candidate substance in order to assess the relative inhibitory capability of the candidate substance.

C. Inhibition of P-TEFb-Viral-Mediated Transcription Elongation

The invention next provides methods for assaying for, or for confirming the identification of, candidate viral transcription inhibitors, based upon the use of viral transcription elongation assays. To identify a viral transcription elongation inhibitor in this manner, one would again 5 conduct parallel or otherwise comparatively controlled transcription elongation assays and identify a compound that inhibits transcription elongation, and preferably, that inhibits transcription elongation of viral transcripts and not human transcripts.

In these assays, one would admix a candidate substance with a transcriptionally 10 competent composition under conditions that would normally result in the generation of elongated human and viral RNA transcripts but for inclusion of an inhibitory substance. One would then identify a positive inhibitory substance as one that prevented or significantly reduced the generation of elongated RNA transcripts upon addition to the assay system. Most preferably, one would identify a positive anti-viral inhibitory substance as one that prevented or significantly reduced the generation of elongated viral RNA transcripts, but that did not prevent or significantly reduce the generation of elongated human RNA transcripts, upon addition to the 15 assay system.

Assays of this type are described herein and are generally performed using effective 20 amounts of human and viral nucleic acid templates, viral transcriptional transactivator proteins, P-TEFb enzyme complex, RNA polymerase II, each four nucleotides, which of course, includes ATP that provides the energy. Preferably, other partially purified nuclear components are included in the assay, as disclosed herein in the Detailed Examples. The productive viral RNA 25 elongation is measured relatively in the presence of the candidate substance, generally after establishing the basal levels of productive viral RNA elongation in the absence of the candidate substance. In this manner, one can assess the relative inhibitory capability of the candidate substance.

D. Second Generation Inhibitors

In addition to the inhibitory compounds initially identified, the inventor also contemplates that other sterically similar compounds may be formulated to mimic the key portions of the structure of the inhibitors. Such compounds, which may include peptidomimetics
5 of peptide inhibitors, may be used in the same manner as the initial inhibitors.

Certain mimetics that mimic elements of protein secondary structure are designed using the rationale that the peptide backbone of proteins exists chiefly to orientate amino acid side chains in such a way as to facilitate molecular interactions. A peptide mimetic is thus designed
10 to permit molecular interactions similar to the natural molecule.

Some successful applications of the peptide mimetic concept have focused on mimetics of β -turns within proteins, which are known to be highly antigenic. Likely β -turn structure within a polypeptide can be predicted by computer-based algorithms, as discussed herein. Once the component amino acids of the turn are determined, mimetics can be constructed to achieve a similar spatial orientation of the essential elements of the amino acid side chains.

The generation of further structural equivalents or mimetics may be achieved by the techniques of modeling and chemical design known to those of skill in the art. The art of computer-based chemical modeling is now well known. Using such methods, a chemical that specifically inhibits viral transcription elongation can be designed, and then synthesized, following the initial identification of a compound that inhibits RNA elongation, but that is not specific or sufficiently specific to inhibit viral RNA elongation in preference to human RNA elongation. It will be understood that all such sterically similar constructs and second generation
25 molecules fall within the scope of the present invention.

E. Methods of Inhibiting P-TEFb

In still further embodiments, the present invention is concerned with methods of inhibiting productive RNA elongation, and preferably, with preferentially inhibiting productive

viral RNA elongation. These methods generally comprise exposing an RNA elongation complex to an effective concentration of a P-TEFb or elongation inhibitor identified in accordance with the candidate screening assay embodiments of the present invention. Where productive viral RNA elongation is to be inhibited, the RNA elongation complex is a complex that comprises 5 viral nucleic acids and transcriptional transactivator proteins, which complex is often located within a virally infected cell.

This aspect of the invention is intended for use in inhibiting the P-TEFb-mediated viral transcription that occurs in various retroviral and other viral infections, such as HIV and HSV 10 infections, and even in adenoviral infections. It is contemplated that the use of such inhibitors to block the transactivating functions of the Tat protein in infected cells of patients suffering with AIDS or HIV-infected states will be useful in itself and/or in conjunction with other anti-viral therapies. Inhibitors designed from dual viral inhibition assay will be particularly useful in the treatment of AIDS because the normal cellular function of P-TEFb should not be influenced. 15

It is also contemplated that P-TEFb will be involved in controlling cellular proliferation. A number of oncogenes are controlled at the elongation phase of transcription and P-TEFb is a key player in this control. The small subunit of human P-TEFb has been localized to a region of 20 a human chromosome that contains a gene involved in cancer. The inventor has also found that P-TEFb levels correlate with the growth rate of cells. Mutations in either subunit of P-TEFb could have profound effects on the cellular proliferation. Gene therapy is therefore contemplated for treatment of conditions that result from P-TEFb mutations. Drugs that affect the cellular function of P-TEFb are contemplated in the treatment of certain types of cancer, irrespective of their specificity for inhibiting certain transcript formation. 25

VII. Pharmaceutical Compositions

A. Pharmaceutically Acceptable Carriers

Aqueous compositions of the present invention comprise an effective amount of one or more of the inhibitors of the invention dissolved or dispersed in a pharmaceutically acceptable carrier or aqueous medium. The phrases "pharmaceutically or pharmacologically acceptable" 30

refer to molecular entities and compositions that do not produce an adverse, allergic or other untoward reaction when administered to an animal, or a human, as appropriate.

As used herein, "pharmaceutically acceptable carrier" includes any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, its use in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions. For human administration, preparations should meet sterility, pyrogenicity, general safety and purity standards as required by FDA Office of Biologics standards.

The biological material should be extensively dialyzed to remove undesired small molecular weight molecules and/or lyophilized for more ready formulation into a desired vehicle, where appropriate. The active compounds will then generally be formulated for parenteral administration, *e.g.*, formulated for injection *via* the intravenous, intramuscular, sub-cutaneous, intralesional, or even intraperitoneal routes. The preparation of an aqueous composition that contains an RNA elongation transcription inhibitor agent as an active component or ingredient will be known to those of skill in the art in light of the present disclosure. Typically, such compositions can be prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for using to prepare solutions or suspensions upon the addition of a liquid prior to injection can also be prepared; and the preparations can also be emulsified.

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions; formulations including sesame oil, peanut oil or aqueous propylene glycol; and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. In all cases the form must be sterile and must be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi.

Solutions of the inhibitory compounds as free base or pharmacologically acceptable salts can be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose. Dispersions can also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations contain a preservative to prevent the growth of microorganisms.

An inhibitor or antagonist of the present invention can be formulated into a composition in a neutral or salt form. Pharmaceutically acceptable salts, include the acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like.

The carrier can also be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and vegetable oils. The proper fluidity can be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. The prevention of the action of microorganisms can be brought about by various antibacterial ad antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions are prepared by incorporating the active inhibitory compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized active ingredients into a sterile vehicle which

contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum-drying and freeze-drying techniques which yield a powder of the active ingredient plus any additional desired ingredient from a previously sterile-
5 filtered solution thereof.

In terms of using peptide inhibitors as active ingredients, the technology of U.S. Patents 4,608,251; 4,601,903; 4,599,231; 4,599,230; 4,596,792; and 4,578,770, each incorporated herein by reference, may be used. The preparation of more, or highly, concentrated solutions for direct
10 injection is also contemplated, where the use of DMSO as solvent is envisioned to result in extremely rapid penetration, delivering high concentrations of the active agents to a small body area.

Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically effective. The formulations are easily administered in a variety of dosage forms, such as the type of injectable solutions described above, but drug release capsules and the like can also be employed.
15

For parenteral administration in an aqueous solution, for example, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous, intramuscular, subcutaneous and intraperitoneal administration. In this connection, sterile aqueous media which can be employed will be known to those of skill in the art in light of the present disclosure. For example, one dosage could be dissolved in 1 ml of isotonic NaCl
20 solution and either added to 1000 ml of hypodermoclysis fluid or injected at the proposed site of infusion, (see for example, "Remington's Pharmaceutical Sciences" 15th Edition, pages 1035-1038 and 1570-1580). Some variation in dosage will necessarily occur depending on the condition of the subject being treated. The person responsible for administration will, in any event, determine the appropriate dose for the individual subject.
25

The active inhibitors or agents may be formulated within a therapeutic mixture to comprise about 0.0001 to 1.0 milligrams, or about 0.001 to 0.1 milligrams, or about 0.1 to 1.0 or even about 10 milligrams per dose or so. Multiple doses can also be administered.

5 In addition to the compounds formulated for parenteral administration, such as intravenous or intramuscular injection, other pharmaceutically acceptable forms include, e.g., tablets or other solids for oral administration; liposomal formulations; time release capsules; and any other form currently used, including cremes.

10 Additional formulations which are suitable for other modes of administration include suppositories. For suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1%-2%.

15 Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders.

20 In certain defined embodiments, oral pharmaceutical compositions will comprise an inert diluent or assimilable edible carrier, or they may be enclosed in hard or soft shell gelatin capsule, or they may be compressed into tablets, or they may be incorporated directly with the food of the diet. For oral therapeutic administration, the active compounds may be incorporated with excipients and used in the form of ingestible tablets, buccal tables, troches, capsules, elixirs, 25 suspensions, syrups, wafers, and the like. Such compositions and preparations should contain at least 0.1% of active compound. The percentage of the compositions and preparations may, of course, be varied and may conveniently be between about 2 to about 75% of the weight of the unit, or preferably between 25-60%. The amount of active compounds in such therapeutically useful compositions is such that a suitable dosage will be obtained.

30

The tablets, troches, pills, capsules and the like may also contain the following: a binder, as gum tragacanth, acacia, cornstarch, or gelatin; excipients, such as dicalcium phosphate; a disintegrating agent, such as corn starch, potato starch, alginic acid and the like; a lubricant, such as magnesium stearate; and a sweetening agent, such as sucrose, lactose or saccharin may be
5 added or a flavoring agent, such as peppermint, oil of wintergreen, or cherry flavoring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar or both. A syrup of elixir may contain the active compounds sucrose as a sweetening agent
10 methyl and propylparabens as preservatives, a dye and flavoring, such as cherry or orange flavor.

It will naturally be understood that suppositories, for example, will not generally be contemplated for use in treating all viral infections. However, in the event that the inhibitors of the invention, or those identified by the screening methods of the present invention, are confirmed as being useful in connection with a particular viral infection of a given tissue or organ, then other routes of administration and pharmaceutical compositions will be more relevant. As such, inhalants, tablets, ophthalmic solutions and other formulations will be appropriate.
15

20 B. Liposomes and Nanocapsules

In certain embodiments, the use of liposomes and/or nanoparticles is contemplated for the introduction of the inhibitors of the invention into host cells. The formation and use of liposomes is generally known to those of skill in the art, and is also described below.
25

Nanocapsules can generally entrap compounds in a stable and reproducible way. To avoid side effects due to intracellular polymeric overloading, such ultrafine particles (sized around 0.1 μm) should be designed using polymers able to be degraded *in vivo*. Biodegradable polyalkyl-cyanoacrylate nanoparticles that meet these requirements are contemplated for use in the present invention, and such particles may be easily made.

Liposomes are formed from phospholipids that are dispersed in an aqueous medium and spontaneously form multilamellar concentric bilayer vesicles (also termed multilamellar vesicles (MLVs). MLVs generally have diameters of from 25 nm to 4 μ m. Sonication of MLVs results
5 in the formation of small unilamellar vesicles (SUVs) with diameters in the range of 200 to 500 \AA , containing an aqueous solution in the core.

The following information may also be utilized in generating liposomal formulations. Phospholipids can form a variety of structures other than liposomes when dispersed in water,
10 depending on the molar ratio of lipid to water. At low ratios the liposome is the preferred structure. The physical characteristics of liposomes depend on pH, ionic strength and the presence of divalent cations. Liposomes can show low permeability to ionic and polar substances, but at elevated temperatures undergo a phase transition which markedly alters their permeability. The phase transition involves a change from a closely packed, ordered structure,
15 known as the gel state, to a loosely packed, less-ordered structure, known as the fluid state. This occurs at a characteristic phase-transition temperature and results in an increase in permeability to ions, sugars and drugs.

Liposomes interact with cells *via* four different mechanisms: Endocytosis by phagocytic cells of the reticuloendothelial system such as macrophages and neutrophils; adsorption to the cell surface, either by nonspecific weak hydrophobic or electrostatic forces, or by specific interactions with cell-surface components; fusion with the plasma cell membrane by insertion of the lipid bilayer of the liposome into the plasma membrane, with simultaneous release of liposomal contents into the cytoplasm; and by transfer of liposomal lipids to cellular or
20 subcellular membranes, or *vice versa*, without any association of the liposome contents. Varying the liposome formulation can alter which mechanism is operative, although more than one may operate at the same time.

C. Kits

Therapeutic kits of the present invention are kits comprising an inhibitor of RNA or viral RNA productive RNA elongation. Such kits will generally contain, in suitable container means, a pharmaceutically acceptable formulation of an inhibitor, or even a gene expressing an inhibitor of a xenobiotic activator of productive RNA elongation in a pharmaceutically acceptable formulation, optionally comprising other anti-viral agents. The kit may have a single container means, or it may have distinct container means for each compound.

When the components of the kit are provided in one or more liquid solutions, the liquid solution is an aqueous solution, with a sterile aqueous solution being particularly preferred. The inhibitors of viral and other xenobiotic activators of productive RNA elongation compositions may also be formulated into a syringeable composition. In which case, the container means may itself be a syringe, pipette, or other such like apparatus, from which the formulation may be applied to an infected area of the body, injected into an animal, or even applied to and mixed with the other components of the kit.

However, the components of the kit may be provided as dried powder(s). When reagents or components are provided as a dry powder, the powder can be reconstituted by the addition of a suitable solvent. It is envisioned that the solvent may also be provided in another container means.

The container means will generally include at least one vial, test tube, flask, bottle, syringe or other container means, into which the inhibitor of viral and other xenobiotic activators of productive RNA elongation are placed, preferably, suitably allocated. Where a second anti-viral therapeutic is provided, the kit will also generally contain a second vial or other container into which this agent may be placed. The kits may also comprise a second/third container means for containing a sterile, pharmaceutically acceptable buffer or other diluent.

The kits of the present invention will also typically include a means for containing the vials in close confinement for commercial sale, such as, e.g., injection or blow-molded plastic containers into which the desired vials are retained.

5 Irrespective of the number or type of containers, the kits of the invention may also comprise, or be packaged with, an instrument for assisting with the injection/administration or placement of the ultimate inhibitor composition within the body of an animal. Such an instrument may be a syringe, pipette, forceps, or any such medically approved delivery vehicle.

10 The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventor to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

15

20

EXAMPLE 1
Purification of P-TEFb, the Limiting Factor in RNA Elongation

25 The reconstruction of DRB-sensitive transcription involves three chromatographically distinct protein fractions, P-TEFa and P-TEFb and factor 2. The present example shows the purification of P-TEFb to near homogeneity and demonstrates that P-TEFb is the only factor strictly required for the transition into productive elongation using an accepted partially fractionated transcription system.

A. Materials and Methods:

1. Materials

30 $[\alpha\text{-}^{32}\text{P}]$ CTP (3000 Ci/mmol) was obtained from ICN (Irvine, CA). Ribonucleoside triphosphates were obtained from Pharmacia-LKB Biotechnology (Piscataway, NJ). DRB

(Sigma, St. Louis, MO) was dissolved in ethanol to 20 mM and stored at -80°C. Phosphocellulose (P-11) and DEAE cellulose (DE-52) were obtained from Whatman (Hillsboro, OR) and prepared according to Price *et al.* (1987). Phenyl Sepharose (Pharmacia) was prepared according to manufacturer's instructions. All other chemicals were reagent grade.

5

2. Chromatography Methods and Initial Fractionation

All chromatography was carried out at 2-4°C. The conductivity of extracts and column fractions were determined with a conductivity meter. A standard curve of conductivity versus 10 mM KCl in HGKEDP (25 mM HEPES, 15% glycerol, indicated concentration of KCl, 0.1 mM EDTA, 1 mM DTT, 0.1% of a saturated solution of phenylmethylsulfonyl fluoride in isopropanol) was used to convert the conductivity measurements into KCl concentration equivalents. Phosphocellulose was obtained from Whatman and was prepared according to the manufacturer's instructions except that the volumes of base and acid washes were reduced from 25 volumes of resin to 5 volumes. Before the resin was stored the KCl concentration was 15 adjusted to 100 mM and the pH was adjusted to 7.6 as measured by a pH meter.

All columns were run in HGKEDP. The K_c cell extract (100 ml at 50 mg/ml protein, see Example 3, Methods, Section A.4 for the preparation of K_c cell extracts) was loaded onto a 500 ml phosphocellulose column at 0.25 column volumes perh. After washing the column with 20 2 column volumes of 100 mM HGKEDP the column was eluted successively with two column volumes each of HGKEDP containing 0.3, 0.4 and 0.75 M HGKEDP. The 0.1 M HGKEDP phosphocellulose flowthrough was brought to 0.3 M and passed through DEAE-cellulose to remove the majority of nucleic acid content before further fractionation. The DEAE flowthrough was loaded onto an 8 ml FPLC Mono Q™ column (Pharmacia) and the proteins were eluted with 25 a 20 column volume gradient from 0.1 to 0.5 M HGKEDP.

The DNase inhibitor, an activity necessary for *in vitro* transcription (*Biochimie* 69:1199-1205, 1987), eluted between 0.29 and 0.31 M HGKEDP and P-TEF α eluted between 0.35 and 0.4 M HGKEDP. The 0.3 M HGKEDP step from phosphocellulose was dialyzed to 0.1 M

HGKEDP and loaded onto a 100 ml DEAE cellulose column. A gradient elution from 0.1 to 0.5 M HGKEDP was performed. TFIIE eluted at 0.12 to 0.15 M HGKEDP, RNA polymerase II at 0.25 to 0.37 HGKEDP, and factor 2 at 0.18 to 0.22 M HGKEDP. The 0.4 M HGKEDP step from phosphocellulose was concentrated using Centricon-30™ concentrators (Amicon) and then 5 dialyzed versus 75 mM HGKEDP for 2 h. The 0.75 M HGKEDP step from phosphocellulose was dialyzed to 0.15 M HGKEDP, and passed through a 100 ml DEAE-cellulose column. This DEAE flowthrough was loaded onto an 8 ml FPLC Mono S™ column equilibrated in 0.15 M HGKEDP. The Mono S™ column was subjected to a gradient elution from 0.15 to 0.5 M HGKEDP. P-TEFb eluted between 0.25 and 0.29 M HGKEDP.

10

3. Purification of P-TEFb

The purification of P-TEFb was carried out twice from 70 to 90 ml of K_cN and once from 155 ml of P-11 0.75 M HGKEDP step generated during the fractionation of *Drosophila* embryonic nuclear extract (obtained from W. Zehring, Wayne State University). Although the exact chromatography steps (described below) differed during the three fractionations, the behavior of the P-TEFb activity from both sources was nearly identical. Two schemes used to purify the P-TEFb used in this work are summarized below.

20 **Scheme 1:** P-TEFb was purified from 90 ml of K_cN (approximately 3.0 g protein). P-TEFb eluted between 0.55 and 0.65 M HGKEDP during gradient elution of a 500 ml P-11 column from 0.15 M to 1.0 M HGKEDP. P-TEFb containing P-11 fractions (250 ml) were adjusted to 0.5 M HGAEDP (HGKEDP with ammonium (NH₄)₂SO₄ substituted for KCl) followed by loading onto a 26 ml Phenyl Sepharose™ column which was eluted with a gradient from 0.5 M to 0 M HGAEDP. P-TEFb eluted from Phenyl Sepharose between 0.12 M and 0 M HGAEDP. P-TEFb containing fractions (23 ml) at 0.06 M HGAEDP (equivalent to 150 mM HGKEDP by conductivity) were pooled and allowed to flow through an 8 ml Mono Q™ equilibrated in 0.15 M HGKEDP directly onto a 1 ml Mono S™ column. The Mono S™ column was then eluted with a gradient from 0.15 M to 0.45 M HGKEDP during which P-TEFb eluted in 2.5 ml (50 mg protein) between 0.25 M and 0.29 M HGKEDP. A 200 µl sample from Mono S™

fraction 30 was loaded onto a 4.25 ml, 18%-35% glycerol gradient with a 500 ml 1M HGKE overlay and centrifuged at 55,000 rpm (287,000 g_{av}) in a Beckman SW 55 Ti rotor at 1°C for 44 h.

5 **Scheme 2:** P-TEFb was purified from 155 ml (265 mg protein) of P11 0.4M to 0.75 M step from embryonic nuclear extract. The initial P-11 step fraction was adjusted to 0.75 M HGAEDP before loading onto a 26 ml Phenyl Sepharose which was then gradient eluted from 0.5 M to 0 M HGAEDP. P-TEFb activity eluted between 0.12 M and 0M HGAEDP in 17 ml and was then dialyzed to 175 mM HGKEDP before being passed through a 5.0 ml DE-52 column. The DE-52 flowthrough (19ml, 3.4 mg protein) was loaded directly onto a 1 ml Mono S™ column which was gradient eluted from 175 mM to 500 mM HGKEDP. P-TEFb activity eluting between 0.25 M and 0.29 M HGKEDP was dialyzed and loaded onto a 1 ml Mono Q™ column at 50 mM HGKEDP, followed by gradient elution from 50 mM to 450 mM HGKEDP. P-TEFb activity was found in both the column flowthrough and early gradient fractions. Both pools of P-TEFb activity were combined and rechromatographed over a 1 ml Mono S™ column loaded at 75 mM and step eluted at 400 mM HGKEDP. P-TEFb eluted in two 0.2 ml fractions. A 125 µl sample from one 0.2 ml fraction was loaded onto a 5 ml, 15%-35% glycerol gradient and centrifuged at 55,000 rpm (287,000 g_{av}) in a Beckman SW 55 Ti rotor at 1°C for 40.5 h.

20 **4. In Vitro Transcription**

Two general types of transcription reactions were performed both using the actin Act5C template (Kephart *et al.*, 1992) linearized with *Hpa I*. Pulse-chase reactions were generally as described previously (Marshall and Price, 1992) and began with a 10 min preincubation (6 ml/reaction) containing 20 mM HEPES, 5 mM MgCl₂, 45-50 mM HGKEDP, 33 mg/ml DNA template and extract or K_c-FT (as described in Example 1, Section B.4). Transcription was initiated by the addition of 2 ml of pulse solution which contained 5 mCi [α -³²P]CTP and brought the reaction to 600 mM in GTP, ATP, and UTP. The true specific activity of the pulse is determined by the contamination of CTP in the other NTPs which the inventor has estimated to be 1 mM. The pulse was continued for 15 seconds after which the reaction was brought to 1.2

mM CTP by the addition of 12 ml of chase solution. The reactions were stopped by the addition of 200 ml of a Sarkosyl solution (1% Sarkosyl, 100 mM NaCl, 100 mM Tris pH 8, 10 mM EDTA and 100 mg/ml tRNA). Sample workup and analysis of labeled transcripts in denaturing gels was as described previously (Price *et al.*, 1987). Briefly, samples were phenol extracted
5 with an equal volume of water saturated phenol. The aqueous phase was precipitated for 10 min at -80°C after addition of 3 volumes of 95% ethanol containing 3 M ammonium acetate. The samples were spun at 15,000 rpm in a microfuge for 10 min. The supernatant was removed and discarded. The pellet and tube were washed with 200 ul of 70% ethanol and spun at 15,000 rpm
10 for 3 min. The pellet was then dried, under vacuum, before being dissolved in gel loading buffer (0.25 × TBE, 8 M urea, with bromophenol blue and xylene cyanol). After heating the samples for 3-5 min at 80°C, they were loaded onto a 6% (or other concentration as desired) polyacrylamide gel cast in 1 × TBE, 6 M urea. After running, the gels were soaked in water containing 1 ug/ml ethidium bromide for 15 min before being dried, under vacuum, and exposed
15 to autoradiographic film.

B. Results

1. Requirement of the CTD for Elongation

Transcription reactions to assay P-TEF activity in partially purified fractions were performed using a continuous labeling protocol, basically as described previously (Price *et al.*,
20 1987). These reactions contained 20 mM HEPES, 5 mM MgCl₂, 600 mM each of GTP, ATP and UTP, 30 mM CTP, 55-60 mM KCl, 3-4 mCi [α -³²P]CTP, 5 mg/ml DNA template, and various protein-containing fractions in a total volume of 12.5 ml. A typical P-TEFb assay contained 0.2 ml of DNase inhibitor (DI), 0.2 ml of RNA Polymerase II, 0.2 ml of dTFIIE, 1.5 ml of concentrated P11-0.4M step fraction, 0.1-0.2 ml of factor 2 and 0.5 ml of P-TEFa. A
25 solution containing buffer, DNA, NTPs and MgCl₂ was added last to start the reactions. Reactions were incubated 20 min at 23°C and stopped as described above.

Using previous work on the chromatography of initiation factors as a guide (Price *et al.*, 1987), the fractionation of K_cN and reconstruction of DRB-sensitive transcription was

accomplished as described above. Complete reconstruction after the first column using flow through (FT), 0.3 M, 0.4 M and 0.75 M steps gives rise to DRB-sensitive transcription. When these four fractions were subjected to further chromatography, a system that only gave rise to DRB-insensitive transcripts was generated. The reconstruction required four fractions: DNase 5 inhibitor (D.I.), TFIIE, RNA polymerase II and a crude P11-0.4 M step. The P11-0.4 M step fraction includes at least TFIIB and TFIIF and probably TFIID and TFIIH.

2. DRB-sensitive Transcription Requires Three Protein Fractions

Besides the fractions required for efficient initiation, three additional fractions are 10 required to reconstruct efficient DRB-sensitive transcription. Two of these fractions contain P-TEFa and P-TEFb. The third is an activity previously described as factor 2 (Price *et al.*, 1987). Using partially purified factors, P-TEFb is the only fraction that is essentially required for elongation. P-TEFa and factor 2 both lead to stimulations in the level of runoff transcript. P-TEFb alone was able to support a very low, but detectable level of DRB-sensitive 15 transcription.

3. P-TEFb Has Two Subunits

Analysis of the final two stages of the purification, chromatography on Mono S™ and glycerol gradient sedimentation, indicate that P-TEFb activity correlates with fractions containing two polypeptides with apparent molecular weights of 124 and 43 kDa. P-TEFb was purified to near homogeneity three times, twice from K_cN and once from *Drosophila* embryonic nuclear extract with nearly identical results. Glycerol gradient analysis of P-TEFb purified using scheme 1 gave nearly identical results. Comparison of the sedimentation of known proteins to that of P-TEFb indicates that the factor is a heterodimer.

25

4. P-TEFb acts after initiation

After purification of P-TEFb a transcription competent nuclear extract depleted of P-TEFb activity was generated. K_cN was passed through P-11 at 0.4 M KCl. The resulting K_c-FT was capable of initiating transcription as efficiently as whole K_cN, but produced only 10% of

the DRB-sensitive transcripts (compare K_cN with K_c-FT alone). Addition of P-TEFb to the K_c-FT restored the ability to generate DRB-sensitive runoff transcripts. The level of DRB-sensitive transcription achieved was dependent upon the concentration of P-TEFb added back. Even with the highest levels of P-TEFb added back, the ratio of DRB-sensitive to abortive transcripts 5 remained similar to that seen previously in extract (Marshall and Price, 1992). Importantly, since P-TEFb was added to the reactions during the chase, the factor must have acted after initiation on the early elongation complexes.

P-TEFb is not required for initiation, does not associate strongly with preinitiation 10 complexes (Marshall and Price, 1992) but is shown here to act during elongation. The timing of DRB sensitivity (Marshall and Price, 1992) and P-TEFb action coincide closely. This is consistent with the activity of P-TEFb being inhibited by DRB. Since DRB is canonically a kinase inhibitor, these data suggest that P-TEFb is a cyclin kinase. Likely targets for a transcription factor kinase that acts early during elongation are RNA polymerase II or a basal 15 initiation factor such as TFIIF. The subunit composition of P-TEFb displays no obvious similarity to known kinases involved in transcription regulation.

EXAMPLE 2

Control of RNA Polymerase II Elongation Potential by P-TEFb

A. Material and Methods

1. Materials

[α-³²P]CTP (3000 Ci/mmol) was obtained from ICN. Ribonucleoside triphosphates were obtained from Pharmacia-LKB Biotechnology. DRB (Sigma) was dissolved in ethanol to 10 20 mM and stored at -80°C. H-8 was obtained from Seikagaku America (Ijamsville, MD) and was dissolved in 20 mM HEPES, pH 7.6 to 20 mM and stored at 4°C. The magnetic concentrator used was the MPC-E from Dynal. All other chemicals were reagent grade.

2. Proteolysis of RNA Polymerase II

Partially purified RNA polymerase II was treated with 0.27 µg/ml chymotrypsin (Sigma) at 27°C in HGED (25 mM HEPES pH 7.6, 15% glycerol, 0.1 mM EDTA, 1 mM DTT) plus 110 mM KCl for times ranging from 0 to 20 min. Digestions were terminated by the addition of 5 trypsin inhibitor (Sigma) to 20 µg/ml. The extent of proteolysis of the RNA polymerase II subunits was assayed by SDS-PAGE followed by silver staining.

3. Proteolysis of Early Elongation Complexes

Preinitiation complexes were formed on an immobilized actin template digested with 10 *HpaII* (780 nt runoff) as described in Marshall and Price (Marshall and Price, 1992). The preinitiation complexes were isolated, washed once with 55 mM HKB (20 mM HEPES, 55 mM KCl and 200 µg/ml BSA) and resuspended into 55 mM HKB. Transcription was initiated by the addition of a pulse solution, which contained 5 µCi of [α -³²P]CTP and brought the reaction mixture to 600 µM in ATP, GTP and UTP and 2 mM MnCl₂. MnCl₂ increased the rate of initiation of preinitiation complexes, thereby, increasing the number of polymerases in early elongation complexes after a short pulse. After 15 seconds the reaction was stopped by the addition of EDTA to 10 mM.

These early elongation complexes were washed 3 times with 1 M HMKB (20 mM HEPES, 5 mM MgCl₂, 1 M KCl and 200 µM BSA) then once with 55 mM HMKB (20 mM HEPES, 5 mM MgCl₂, 55 mM KCl and 200 µM BSA). The washed early elongation complexes were resuspended in 55 mM HMKB and incubated with the indicated amount of chymotrypsin for 10 min. Proteolysis was terminated by adding trypsin inhibitor to 0.1 mg/ml. After concentration, digested early elongation complexes were either resuspended into 55 mM HMKB 25 and chased with 600 µM of each NTP for 10 min in the presence or absence of K_c nuclear extract and 0.1 µl P-TEFb, or analyzed by SDS-PAGE followed by immunoblotting with affinity purified RNA polymerase II antibody.

The P-TEFb used in FIG. 2A, FIG. 2B and FIG.3 was purified from *Drosophila* embryonic nuclear extract as described in Example 1 with one addition. Material eluting from the Phenyl-Sepharose column was loaded directly onto a 10.0 ml ceramic Hydroxyl-apatite column (Bio-Rad CHT10). The column was then eluted with a linear gradient of potassium phosphate from
5 10 mM to 750 mM in 25 mM HEPES, pH 7.6, 15% glycerol. P-TEFb eluted between 400 mM and 500 mM phosphate. Pooled fractions containing P-TEFb were then dialyzed and chromatographed on Mono S. P-TEFb eluting from Mono S™ still had significant nucleic acid contamination so the material was subjected to chromatography on Mono Q™ followed by Mono S™ for re-concentration. The peak fraction from the final Mono S™ column contained about 0.5
10 mg/ml P-TEFb.

7. CTD Kinase Assay

0.4 μ l of purified RNA polymerase II and various protein samples were mixed in 18 μ l of 55 mM HMK. The reaction was then initiated by the addition of 2 μ l of a solution containing 2 μ Ci of [γ -³²P]ATP (ICN) and unlabeled ATP at 10 μ M, or only unlabeled ATP at concentrations from 1 to 100 μ M or other NTPs or dNTPs at concentrations of 1 to 100 μ M. Reactions were incubated for the indicated times at 23°C and then terminated with SDS loading buffer. Samples were analyzed on a 6-15% SDS polyacrylamide gel which was silver stained, dried and subjected to autoradiography if the assay contained label.

B. Results

1. Productive Elongation Requires the CTD

Drosophila RNA polymerase II was treated with chymotrypsin for increasing times to gradually remove the CTD (FIG. 1). Trypsin inhibitor was added to aliquots of the digestion reaction after 0, 2, 8, or 20 min. When trypsin inhibitor was added to a similar reaction before the chymotrypsin, no digestion took place during a subsequent 20 min incubation indicating that the protease was inactivated. Intact or truncated forms of the polymerase were then used to drive transcription from the Act5C promoter using fractions derived from *Drosophila* K_c cell nuclear extract containing factors needed for initiation and productive elongation.

Most of the RNA polymerase II was removed by the fractionation procedure; however, some runoff transcripts were detected in the absence of added RNA polymerase II. When intact RNA polymerase II was added the runoff signal increased dramatically and was sensitive to
5 DRB. The amount of DRB-sensitive, runoff transcript decreased as the CTD was removed. The amount of shorter, DRB-insensitive, abortive transcripts increased with added RNA polymerase II but did not change in amount as the CTD was removed. Quantitation of the autoradiograph and protein gel showed that the level of runoff transcript was directly related to the amount of polymerase containing the CTD while the generation of abortive transcripts was unaffected by
10 loss of the CTD (FIG. 1).

Earlier findings, using a minimal set of *Drosophila* fractions required for initiation, indicated there was no effect when substituting CTD-less polymerase for intact polymerase (Zehring *et al.*, 1988). These earlier results were obtained using fractions that did not contain factors needed for the generation of DRB-sensitive productive elongation complexes (Marshall and Price, 1995). Therefore, the negative effect of removal of the CTD indicates that the CTD is involved in the transition into productive elongation.
15

To address the possibility that truncation of the CTD in the previous study had an effect on initiation that resulted in the formation of exclusively DRB-insensitive complexes, a protocol for the truncation of the CTD after initiation was developed. Early elongation complexes were formed on an immobilized template and then washed with buffer containing 1 M KCl which removes uninitiated RNA polymerase II. The proteins found in early elongation complexes were stripped from the beads with SDS and analyzed by SDS-PAGE followed by western blotting.
20
25 Rabbit anti-*Drosophila* large subunit antibodies to a non-CTD containing domain of the large subunit of RNA polymerase II were generated (see above in Methods, Section 4) and used to probe the western blot.

Treatment with increasing amounts of chymotrypsin resulted in removal of the CTD from
30 RNA polymerase II in early elongation complexes. Essentially complete truncation occurred

when 0.2 µg/ml or more of chymotrypsin was used. The protease treatment did not negatively affect the ability of the isolated early elongation complexes to continue elongation during the subsequent chase. Except for the increase in the longest transcripts at the higher protease concentrations, the typical pattern of transcripts seen during abortive elongation on the actin 5 template (Marshall and Price, 1992; Xie and Price, 1996) were detected. The increase in the longest transcripts is due to the removal of the remaining trace amount of factor 2 (Price *et al.*, 1987) which normally exerts a negative effect on elongation (Xie and Price, 1996).

To assess the ability of the polymerase in early elongation complexes to enter productive 10 elongation, nuclear extract was added back with the chase. In this study the extracts were complemented with a constant amount of additional P-TEFb which is normally limiting in the extracts. DRB-sensitive runoff transcripts due to P-TEF action were visible in the western blot lanes using non-proteolyses complexes. As increasing amounts of chymotrypsin were used, the 15 early elongation complexes lost the ability to form long DRB-sensitive transcripts. The long transcripts seen at high protease levels in the early elongation complexes alone were not seen when K_cN was added because of the effect of endogenous factor 2 in the extract (Xie and Price, 1996). These results show that the CTD is required for the generation of long DRB-sensitive transcripts.

20 2. P-TEFb is a CTD Kinase

The requirement of the CTD for the generation of DRB-sensitive long transcripts and the inhibition of the process by the kinase inhibitor DRB (Marshall and Price, 1992; Marshall and Price, 1995) prompted studies to determine if P-TEFb was a CTD kinase. Studies similar to those in Example 1, and prepared in the same manner, showed that incubation of P-TEFb with 25 intact RNA polymerase II caused the incorporation of phosphate into the large subunit and a shift to the I_{Io} form.

Truncation of the CTD by chymotrypsin resulted in a loss of the ability of P-TEFb to phosphorylate the polymerase. The polymerase that was treated with chymotrypsin for 0, 2, 8, or

20 min followed by addition of trypsin inhibitor, see Section B.1, was used as a substrate for pure *Drosophila* P-TEFb. Using the incorporation of label into the large subunit of RNA polymerase II as an assay (see Methods, Section A.7) the ability of P-TEFb to label the large subunit was directly related to the amount of the CTD remaining on the subunit. The intact 5 polymerase was labeled very efficiently, while the CTD-less polymerase (20 min digestion) was not detectably phosphorylated. Intermediate truncations gave rise to intermediate levels of phosphorylation.

To correlate the CTD kinase activity with P-TEFb function in transcription, fractions 10 from a gradient elution of P-TEFb from a Mono S™ column and a subsequent glycerol gradient were tested in both assays. P-TEFb was about 25% pure in the peak fraction (Peterson *et al.*, 1992) from Mono S™ (Example 1; Marshall and Price, 1995). CTD kinase activity coeluted with 15 the transcription activity of P-TEFb on Mono S™. Fractions from the glycerol gradient analysis of P-TEFb from Mono S™ fraction 30 were analyzed for transcription and CTD kinase activities on a silver-stained 6-15% SDS-polyacrylamide gel (SDS-PAGE). Both activities again co-migrated with each other and with the 124 and 43 kDa subunits of P-TEFb previously identified (Example 1; Marshall and Price, 1995). P-TEFb subunits, transcription function and CTD kinase activity correlated across all columns assayed.

20 3. Characterization of P-TEFb CTD Kinase

To begin to establish the basic parameters of the CTD kinase activity of P-TEFb, P-TEFb 25 was incubated for 5 min with RNA polymerase II and various cold nucleotides. The products were subjected to SDS-PAGE and silver stained to visualize the RNA polymerase II. The mobility shift from the unphosphorylated IIa form to the highly phosphorylated IIo form of the large subunit was used to ascertain kinase activity. Partial shifts indicated lower levels of incorporation of phosphate. P-TEFb was capable of utilizing the purine nucleotides in rough order of efficiency ATP >> dATP = GTP > dGTP. ATP was used 10-100 times more efficiently than dATP. GTP was used ~2-fold more efficiently than dGTP. P-TEFb was unable to utilize any of the pyrimidine nucleotides at 100 µM. In a similar study it was found that even at high

concentration, *i.e.* 600 μ M CTP, 600 μ M UTP, 600 μ M dCTP, or 600 μ M dTTP, other nucleotides were not used as substrates for P-TEFb in the phosphorylation of RNA polymerase II.

5

4. Kinetics of Phosphorylation

The kinetics of phosphorylation of the CTD were examined with three concentrations of P-TEFb and 10 μ M ATP. At all concentrations of P-TEFb intermediates between II a and II o were seen at early time points indicating a progressive increase in phosphorylation. At the lowest concentration of P-TEFb ~50% of the large subunit molecules were progressively phosphorylated, even though the molar ratio of P-TEFb to RNA polymerase II was about 1:50. This result indicates that P-TEFb was not stably associated with the polymerase during the addition of all phosphates and, therefore, was not completely processive. As the concentration of P-TEFb was increased the fraction of the large subunit that was shifted to II o increased to about 90%.

10

15

20

25

Comparison of the kinetics of the appearance of the intermediate forms indicates that increasing the P-TEFb concentration had a modest but non-linear effect on the rate of the phosphorylation. For example, comparing 1 min with 0.01 μ l of P-TEFb to 30 seconds with 0.05 μ l indicates that although less large subunit was affected, an equivalent mobility shift was obtained. At all P-TEFb concentrations tested there were at least some polymerase molecules that received no or very few phosphates while others were heavily phosphorylated to the II o form. These data are consistent with a slow initial phosphorylation event(s) followed by more rapid subsequent phosphate incorporation. The different rates could be due to conformational changes in the polymerase or CTD required for initial kinase action or to increased reactivity of partially phosphorylated CTD.

5. Sensitivity of the Kinase to Inhibitors

The sensitivity of the kinase to both DRB and H-8 under various conditions was observed. At 10 μ M ATP (the same concentration of ATP used in the labeling assay), P-TEFb

was strongly inhibited by DRB at 10 μ M. Under the same conditions, approximately 10 times as much H-8 was required to inhibit the kinase to the same level. This 10-fold difference was seen consistently under all conditions tested. At 100 μ M ATP both drugs were less effective. The inhibition by DRB and H-8 was directly related to the concentration of nucleotides used.

5 Because the inhibition by DRB and H-8 was affected by nucleotide levels the relative inhibition between DRB and H-8 can be utilized as a more useful way to compare P-TEFb with other kinases.

Since the CTD kinase activity of P-TEFb was sensitive to H-8, the effect of the drug during transcription in nuclear extracts was examined. Increasing amounts of DRB or H-8 were included in a continuous labeling assay. As seen earlier runoff transcription was very sensitive to DRB with a 50% inhibition point of 0.7 μ M (FIG. 2A). Under identical transcription conditions, H-8 had a 50% inhibition point of 7 μ M (FIG. 2B). This 10-fold difference is the same as that seen with the CTD kinase assay. As expected for an inhibitor of productive elongation, initiation and abortive transcription were unaffected by H-8 even at concentrations that severely inhibited the appearance of runoff. These data show that P-TEFb is the target of these kinase inhibitors during transcription.

6. P-TEFb Phosphorylates the CTD of RNA Polymerase II in Early Elongation

The data herein show that an intact CTD is required for productive elongation and that P-TEFb can phosphorylate pure RNA polymerase II. Since P-TEFb acts during elongation (Marshall and Price, 1995), it should phosphorylate the CTD if the polymerase is in an early elongation complex. To determine if the polymerase is in an early elongation complex western blot analysis was used to determine the phosphorylation state of the polymerase in isolated transcription complexes.

Preinitiation complexes were formed on an immobilized actin template as described above in Methods, Section 3. When these complexes were washed with low salt buffer they remained intact and western blot analysis indicated that the large subunit of RNA polymerase II

was hypophosphorylated. Incubation of the preinitiation complexes with ATP caused a significant shift of the polymerase to the IIo form. The antibodies used to detect the large subunit of RNA polymerase II reacted more strongly with the IIa form of the large subunit than either the phosphorylated IIo form or the truncated IIb form. This phosphorylation was
5 unaffected by the presence of 20 μ M DRB.

This result is consistent with the previous data that P-TEFb is not associated with preinitiation complexes (Marshall and Price, 1992; Marshall and Price, 1995; see Example 1). Essentially all of the RNA polymerase II was removed by washing with buffer containing 1 M
10 KCl. When the preinitiation complexes were incubated under transcription conditions for a short time, a portion of the polymerases initiated and, therefore, remained associated with the template during the high salt wash. The CTD kinase found in the preinitiation complex was no longer associated with the polymerase because incubation with ATP had no effect. However when increasing amounts of P-TEFb were incubated with the high salt washed early elongation complexes, the IIa form of the polymerase decreased and the IIo form increased. The ability of
15 P-TEFb to phosphorylate RNA polymerase II in an early elongation complex was inhibited by 20 μ M DRB.

7. TFIIH Does Not Functionally Substitute for P-TEFb

To determine whether P-TEFb was related to the protein kinase associated with TFIIH (Roy *et al.*, 1994; Feaver *et al.*, 1994; Cismowski *et al.*, 1995; Serizawa *et al.*, 1995; Shiekhattar *et al.*, 1995), P-TEFb was compared with *Drosophila* TFIIH purified from *Drosophila* embryos by the method of Austin and Biggin (1996).

The two proteins did not have any subunits in common when analyzed by SDS-PAGE and silver staining. The large subunit of P-TEFb ran as a doublet which may be due to phosphorylation, since autophosphorylation of both P-TEFb subunits had been previously observed.
25

To compare the activities of the two factors, both were tested in the CTD kinase assay under identical conditions. Using equal concentrations of the two kinases, as determined by relative protein staining, both were able to incorporate similar amounts of $^{32}\text{PO}_4$ into the CTD of RNA polymerase II and cause the shift to the I_{Io} form. The CTD kinase activity of P-TEFb was
5 sensitive to 20 μM DRB while that of TFIIH was unaffected. Since DRB inhibition could be affected by enzyme levels titration studies were used to determine that the levels of P-TEFb and TFIIH were not saturating. To determine if TFIIH could functionally replace P-TEFb similar amounts of both proteins were tested in the P-TEFb -dependent K_c-FT (Marshall and Price, 1995;
10 Example 1) transcription assay. As expected, addition of increasing amounts of P-TEFb led to an increase in DRB-sensitive runoff transcripts (FIG. 3). In contrast, addition of even the highest levels of TFIIH had little effect on the amount of runoff even though the added TFIIH had high CTD kinase activity when assayed using purified RNA polymerase II. Therefore, it appears that TFIIH cannot substitute for P-TEFb during the transition into productive elongation.

The 10-fold difference between levels of DRB and H-8 required to achieve equal inhibition of CTD phosphorylation remained constant under all conditions examined. Since a 10-fold difference in the levels of DRB and H-8 was required to inhibit transcription, this indicates that the effect of the drugs in transcription is due to inhibition of the CTD kinase activity of P-TEFb.
15

After the initial transition into productive elongation has been passed, it may be necessary to maintain the highly phosphorylated state of the polymerase. However, the results of Egyhazi *et al.* (1996) and Kephart *et al.* (1992) suggest that if CTD kinases are needed for maintenance of the I_{Io} form during elongation then these kinases are not sensitive to DRB. Therefore, P-TEFb
20 does not appear to be involved in maintenance. Maintenance could be accomplished by the inhibition of CTD phosphatase (Chambers *et al.*, 1995) activity.
25

EXAMPLE 3
Drosophila P-TEFb Increases the Processivity of Human RNA Polymerase II

A. Materials and Methods

5 **1. Materials**

Ribonucleoside triphosphates were obtained from Pharmacia LKB Biotechnology Inc. [α-³²P]CTP (3000 Ci/mmol) was purchased from ICN pharmaceuticals. Streptavidin-coated paramagnetic beads (Dynabead™ M280) were obtained from Dynal (Great Neck, N.Y.). All other reagents were as described previously (Marshall and Price, 1992).

10

2. Growth mediums

D22 Medium for Growth of K_c cells (for 50 L):

A:	MgCl ₂ · 6H ₂ O	50g
	MgSO ₄ (Anhydr.)	82g
	CaCl ₂ (Anhydr.)	40.5g
	L-glutamic acid	530g
	Glycine	270g
B:	Potassium hydroxide (45%)	152.5ml
	Sodium hydroxide (50%)	180.5ml
C:	L-malic acid	30.5g
	Succinic acid	2.75g
	Glucose	91g
25	Sodium acetate	0.7g
	Lactalbumin hydrolysate	682g
	Yeastolate	68g
D:	Grace's vitamins, 100ml	
30	E:	Potassium hydroxide (45%) to pH 6.7
	F:	NaH ₂ PO ₄ · H ₂ O (100X) pH 6.7, 500ml (100X = 3.8g/100ml)

The ingredients are added in sequential order (A, then B, then C, then D, then E, then F) and brought to a final volume of 50 L. The medium is filter sterilized by starting with a large-pore filter and ending with a 0.22mm final filter.

5

Grace's Vitamins (modified) for 1 liter:

	<u>chemical</u>	<u>mg/liter</u>
	Biotin	50
10	D-Ca Pantothenate	10
	Choline Chloride	1000
	Folic Acid	10
	i-Inositol	100
	Niacin	100
15	para-aminobenzoic acid	100
	pyridoxine HCl	10
	Riboflavin	10
	Thiamine HCl	10

Add 45% KOH until all material is dissolved. Filter or store at -20°C.

3. Solutions

<u>HGE</u>	<u>4 liters</u>	<u>1 M HGKE</u>	<u>2 liters</u>
25 mM HEPES	23.8 g	25 mM HEPES	11.9 g
15% glycerol	600 ml	15% glycerol	300 ml
0.1 mM EDTA (adjust pH to 7.6 with 50% NaOH)	0.8 ml (0.5 M)	0.1 mM EDTA	0.4 ml (0.5 M)
		1 M KCl	149.1 g
		(adjust pH to 7.6 with 50% NaOH)	

	<u>Buffer A</u>	<u>500 ml</u>	<u>Buffer B</u>	<u>200 ml</u>
30	15 mM KCl	7.5 ml 1 M	1 M KCl	14.9 g
	10 mM HEPES	10 ml (0.5 M)	50 mM HEPES	20 ml (0.5 M)
	2 mM MgCl ₂	1 ml (1 M)	30 mM MgCl ₂	6 ml (1 M)
35	0.1 mM EDTA	0.1 ml (0.5 M)	0.1 mM EDTA	0.04 ml (0.5 M)

Stock solutions: 0.5 M HEPES, pH 7.8; 0.5 M, EDTA, pH 8.3; 1 M MgCl₂.

Complete the above buffers just before use with: 1 mM DTT from 1 M stock (stored at -80°C) and 1:1000 dilution of isopropanol saturated with PMSF at room temperature.

Stock solution: 4M (NH₄)SO₄, pH 7.9

5

All buffers are filter sterilized. All glassware is autoclaved.

4. DNA templates

The *Drosophila* actin A2 (RI/*PstI*) DNA template has been described (Marshall and Price, 1992). Transcription of this template after digestion with *HpaI* yields a 520-nucleotide (nt) runoff transcript. The HIV-1 template (pLTR-4/CAT) derived from HIV-SF2, was obtained from P. Luciw (Sanchez-Pescador *et al.*, 1985). It contains the HIV-1 LTR from -153 to +80 and has the CAT gene downstream of the LTR. Transcription of this construct after digestion with *NcoI* produces a 633 nucleotide runoff transcript. Transcription of the HIV-1 construct after digestion with *BamHI* produces a 1640 nucleotide runoff and digestion with *XbaI* produces a 4000 nucleotide runoff transcript. Biotinylated *Drosophila* A2 DNA templates were made as previously described (Marshall and Price, 1992).

5. K_c Cells and Cell Extracts

Growth of K_c cells and preparation of nuclear extracts are as described (Price *et al.*, 1987).

a. freezing and thawing cells:

Cells are grown in spinner culture to 4×10^6 cells/ml then gently spun in 50 ml conical tubes in a JS-4.2 rotor (Beckman) at 500 rpm/10 min/4°C. Spun cells are resuspended in 4.5 ml ice-cold fetal calf serum and 0.5 ml DMSO or 1/10 original volume. Aliquots of 1 ml each are placed into cryotubes while on ice, frozen at -20°C for 1 h and then transferred to -80°C overnight. Cells can be stored for months at -80°C or longer in liquid N₂ without impairment to function.

Frozen cells are thawed at room temperature and diluted 10-20 fold with D22 media in a 25 ml tissue culture flask. The D22 media is removed and replaced with 6-fold fresh D22 media when cells have attached to the flask (approximately 30 min). Fresh D22 media is added over the next two days until the cells reach the desired density and can be split.

5

b. continuation of line:

The cells should be continued in flasks with care being taken never to allow them to become overgrown, as evidenced by the release of all cells from the bottom of the flask with only gentle tilting. The cells will tolerate lower densities in flasks due to the close proximity of other settled cells, but the cultures should be split at least every 3.5 days or one day before cells become detached by tilting.

10
15
20

c. growth for production of extracts:

Careful control of the conditions of growth of K_c cells is critical for the production of high quality extracts for *in vitro* transcription. The cells must be maintained between 1.5×10^6 and 8×10^6 cells/ml in spinner cultures, preferably between 2×10^6 and 6×10^6 cells/ml. Cells will double every 16 to 18 h when growing well. Cells that are diluted too much will not grow, and cells that reach densities of over 8×10^6 cells/ml will not yield good extracts even if their growth rate is good after dilution and further growth. Extracts should be made when cells are between 3×10^6 and 5×10^6 cells/ml. Most of the cells in a good culture will be attached to another cell. However, the stirring rate should be increased if many cells are attached in chains of four or longer.

d. Preparation of K_c Cell Nuclear Extract, K_cN:

25 This protocol details the production of a nuclear extract from *Drosophila* K_c cells. It has been described in Price, *et al.* (1987). A standard preparation starts with 18 liters of cells, but the protocol can be scaled down or up as desired. Provisions are given to save the cytoplasmic portion of the cells which may contain other cellular factors of interest. Phenol extraction of the cytoplasm will produce massive quantities of high quality K_c cell cytoplasmic RNA. This

procedure produces extracts that are more active than any other extracts reported. The procedure can also be used on other types of cells including mammalian cells. K_c cells are grown in D22 media to a density of 3 to 5 × 10⁶ cells/ml. Prepare, at 4°C, solutions A, B and A + 1/15B. Add DTT and PMSF just prior to use.

5

e. Cell harvesting:

K_c cells are aliquoted into centrifuge tubes, about 65 mls per bottle, and spun in a Beckman JS-4.2 rotor at 4.2K rpm (4000 x g) for 7 min. The supernatant is removed and the pelleted cells are retained. 10 ml Buffer A+1/15 B is added to each centrifuge bottle and the 10 cells are resuspended and transferred to four 40 ml polycarbonate tubes. A 10 ml wash of Buffer A+1/15 B is used to wash all of the bottles and the wash is added to the appropriate polycarbonate tube. The cells are spun in the Beckman JS-13.1 swinging bucket rotor for 5 min at 4,000 rpm (4000 x g).

The supernatant is discarded and the pelleted cells are gently resuspended the cells in Buffer A (10 ml/tube). Resuspended cells are spun a JS-13.1 rotor for 5 min at 5,000 rpm. The slightly cloudy supernatant is discarded. The pelleted cells are resuspended in Buffer A to 40 ml total volume, transferred to a 40 ml Dounce homogenizer and homogenized until cell lysis is greater than 90%. Cell lysis is normally 95%-98%. 3 mls of Buffer B is added and **lyzed** cells are briefly homogenized again. Lysed cells are transferred to a polycarbonate tube.

The lysed cells are spun in a JS-13.1 rotor for 8 min at 8,000 rpm to pellet nuclei. The supernatant (K_cC) is transferred to disposable 50 ml conical tube(s) and placed on ice. 10 mls of Buffer A is added to each tube and nuclei (K_cN) are gently mixed until completely resuspended. 25 The volume of the resuspended nuclei is brought to total volume of 40 ml with Buffer A. The nuclei are uniformly resuspended by gentle homogenization.

The nuclear suspension is transferred to 28 ml Oak Ridge Tubes, 20 ml/tube, 2 ml of 4 M (NH₄)SO₄ is added to each tube (360 mM (NH₄)SO₄ final) and mixed for 30minutes at 4°C.

The resulting suspension is transferred to fresh Oak Ridge tubes. Both the nuclear suspension and the saved supernatant are spun in a Beckman 55.2Ti rotor for 45 min at 45,000 rpm (150,000 x g) to pellet the chromatin.

5 The resulting supernatant from the nuclear suspension is transferred into a fresh Oak Ridge tube containing 5.5g solid (NH₄)SO₄ (0.25g/ml) and mixed for 20 min at 4°C. The spun K_cC supernatant is frozen at -80°C. The nuclear suspension is precipitated with ammonium sulfate in the Beckman 55.2Ti rotor for 15 min at 45,000 rpm and the supernatant is removed.

10 **f. Pellet resuspension and dialysis:**

15 The protein pellet is dissolved in 1/2 of 1 nuclear volume of HGEDP (HGE + DTT and PMSF) and transferred to a 7 ml Dounce homogenizer and homogenize gently until the protein is completely dissolved. 25 ml HGKE is added to 475 ml HGEDP to make 500 ml HGKEDP. the solubilized protein is dialyzed against 500 ml of 50 mM HGKEDP for 2.5-3.0 h. at 4°C. The final salt concentration should be 120-150 mM KCl. Dialized protein is aliquoted and frozen at -80°C.

20 **6. Preparation of HeLa Cell Nuclear Extract**

25 The growth of HeLa cells and preparation of nuclear extracts (NE's) are essentially as described by Dignam *et al.* (1983) with some modification as described below. P-TEFb was purified from *Drosophila* embryonic nuclear extract as previously described in Example 1. This protocol details the production of a nuclear extract from mammalian cervical cancer, HeLa, cells. It is very similar to the K_c cell extract described above in Section A.4 and by Price, *et al.* (1987). Modifications to the cell harvest are described below. All other methods for preparation of HeLa cells are described in Section A.4 above.

a. cell harvest:

HeLa cells are grown on plates or in spinner flasks (5×10^6 - 7.5×10^6 cells/ml) in DMEM/10% Calf Serum. The HeLa cells are either scraped off of T-150 Plates or cells grown in

D22 media are spun down in a Beckman JS-4.2 rotor at 4.2K rpm (4000 x g) for 5 min and the pelleted cells are recovered. Cells are resuspended as before in Buffer A+ 1/15 B except that only ½ of the volume indicated in Section A.4 is used. Volumes of all following isolation and purification steps are modified accordingly.

5

7. Transcription Reactions

A pulse-chase protocol was used in which the template DNA was preincubated with the extract, 20 mM HEPES, pH 7.6 and 10 mM MgCl₂ for 20 min at 30°C. Nucleotides, including [α-³²P]CTP were added to start the pulse; 2 min later, excess cold CTP was added to initiate the chase. Reactions were stopped after the indicated chase times. During the pulse, 10 µl reaction mixtures contained the following: 20 mM HEPES (pH 7.6), 10 mM MgCl₂, 600 µM each GTP, UTP and ATP, 5 µCi of [α-³²P]CTP (~1 µM CTP), 66 mM KCl, 10-40 µg/ml DNA template, and 3 µl of HeLa nuclear extract. For the chase, unlabeled CTP was added to bring the total concentration of CTP to 1.2 mM and the final reaction volume to 12 µl. Reaction mixtures containing 250 mM KCl, were supplemented with KCl at the beginning of the chase. *Drosophila* P-TEFb was added to the preincubation mixture where indicated. Reactions were stopped by adding 200 µl of stop solution (1% Sarkosyl, 50 mM Tris, pH 8.0, 50 mM EDTA, 100 mM NaCl and 100 µg/ml tRNA). The reaction mixtures were phenol extracted and the nucleic acids were ethanol precipitated, washed with 70% ethanol, dried and analyzed by gel electrophoresis. Transcription reactions with K_c cell NE were as described (Kephart *et al.*, 1992).

8. Gel Electrophoresis

For polyacrylamide gel electrophoresis, samples were resuspended in 7 ml of 0.25X TBE (1X TBE contains 89 mM Tris base, 89 mM Boric acid and 2 mM EDTA) and 8 M urea, heated for 3 min at 85°C and analyzed in 6% acrylamide-8 M urea-1X TBE gels. Denaturing agarose gel electrophoresis was performed by resuspending samples in 50% formamide-2.2 M formaldehyde-1X morpholinepropanesulfonic acid (MOPS) buffer (20 mM MOPS, pH 7.0; 5 mM sodium acetate, 1 mM EDTA) and heating the mixture at 60°C for 5 min. Samples were resolved in a 2% agarose gel containing 2.2 M formaldehyde and 1X MOPS buffer.

B. Results

Previous results using *Drosophila* DNA templates and K_c cell nuclear extract (NE) have shown that two classes of RNA polymerase II elongation complexes are present during transcription, *in vitro* (Marshall and Price, 1992). These are either elongation complexes that produce short transcripts or productive elongation complexes capable of making long runoff transcripts. P-TEFb, a factor required for the transition into productive elongation (Marshall and Price, 1995), functions *via* its ability to phosphorylate the CTD of the large subunit of RNA polymerase II in early elongation. Both the function of P-TEFb and its CTD kinase activity are inhibited by low levels of DRB (see Example 2). The present invention demonstrates that a human P-TEFb homolog is involved in controlling elongation in the HeLa transcription system.

1. DRB-Sensitivity of Human RNA Polymerase II Elongation Complexes

To begin to examine elongation control in a HeLa *in vitro* transcription system, the general properties of human elongation complexes were studied using conditions similar to those used to study *Drosophila* complexes in Example 2.

Using a HeLa NE pulse-chase reactions were performed at either 10 or 40 µg/ml template with or without DRB. After a 2 min pulse under limiting CTP conditions the incomplete transcripts had a distinct pattern. More early elongation complexes (EECs) were formed at the higher template concentration as indicated by the increase in intensity of the transcripts after the pulse. Addition of 50 µM DRB inhibited the formation of the longest transcripts during the pulse. When the EECs formed during the pulse were chased for the indicated times, a very processive, DRB-sensitive elongation complex was identified. Even though more polymerase molecules initiated at the higher template concentration, a similar number ultimately reached runoff and the runoff accumulated more quickly at the lower template concentration.

These results indicate that, as has been seen in a *Drosophila* transcription system, P-TEF was limiting for the generation of DRB-sensitive runoff transcripts (Marshall and Price, 1992).

A difference between the properties of the elongation complexes formed in the HeLa system and the *Drosophila* system was that at long time points only 60% (instead of greater than 95%) of the runoff generated was DRB-sensitive.

5 To determine if the DRB-insensitive elongation complexes that were capable of making long transcripts arose from incomplete inhibition by DRB, DRB was titrated from 10 to 135 μ M. Maximum inhibition by DRB occurred at 30 μ M. Increasing the DRB concentration to 135 μ M did not inhibit the formation of the DRB-insensitive processive elongation complexes. Therefore, 50 μ M DRB was used in all subsequent studies. These results show that the HeLa
10 transcription system gave rise to both types of elongation complexes found in the *Drosophila* system and in addition produced a class of DRB-insensitive, processive elongation complexes not seen in the *Drosophila* system.

15 It was possible that the apparent difference between the two systems was merely due to properties of HeLa RNA polymerase II and that the DRB-insensitive transcripts generated by the human polymerase were just longer than those seen in the *Drosophila* system. If this were so, then longer transcripts would be more sensitive to DRB than shorter ones.

20 To determine if this were the case, templates that produced longer runoff transcripts were used. Two min pulse-chase time course assays using DNA templates that generate 1640 or 4000 nucleotide runoff were carried out. At longer time points both DRB-sensitive and DRB-insensitive processive elongation complexes generated 1640 and 4000 nucleotide runoff transcripts. Addition of 250 mM KCl to the chase caused the EEC's to elongate at a slower rate, compared to 60 mM KCl in a normal chase.

25

As was found in the *Drosophila* system (Kephart *et al.*, 1992; Marshall and Price, 1992) high salt suppressed both positive and negative elongation factors. The rate of elongation was slowed when compared to that seen for productive elongation complexes under normal salt conditions. As in the *Drosophila* system (Kephart *et al.*, 1992; Marshall and Price, 1992), early

blocks to elongation (promoter proximal pausing) are apparently removed by the high salt treatment. *Rous sarcoma* virus LTR, adenovirus major late promoter and SV40 promoter constructs were also examined and gave similar results to those seen with the HIV-1 DNA template.

5

2. Demonstration of P-TEFb -like activity in HeLa nuclear extracts

In the *Drosophila* system the DRB-sensitivity of the generation of highly processive elongation complexes is due to the CTD kinase activity of P-TEFb (Marshall and Price, 1995). To examine the similarity of the DRB-sensitive process in HeLa transcription to that of the well characterized *Drosophila* system, the sensitivity of HeLa elongation complexes to both DRB and H-8 were compared.

Two min pulse/45 sec chase transcription assays were performed with variable amounts of DRB or H8. The inhibition of the appearance of runoff was quantitated and plotted versus the concentration of inhibitor (FIG. 4). The 50% inhibition point occurred at 1 μ M DRB compared to 20 μ M H8 (FIG. 4). Greater sensitivity to DRB was also seen with the *Drosophila* system which was 50% inhibited by 0.7 μ M DRB compared to 7 μ M H8 (see Example 2). It has been shown that the sensitivity of the CTD kinase activity of P-TEFb is very dependent on the conditions used in the assay (Marshall and Price, 1995; see Example 1). Conditions used here were similar to those used in the *Drosophila* study. These inhibitor studies indicate that there is a HeLa equivalent to *Drosophila* P-TEFb.

Since P-TEF was first identified using an immobilized template assay, the ability of HeLa NE to generate DRB-sensitive elongation complexes when added back to isolated EECs was examined. *Drosophila* actin A2 DNA template and K_c cell NE was used to form EECs containing *Drosophila* RNA polymerase II. An immobilized A2 template was preincubated with K_c cell NE and then pulsed for 30 sec with [α -³²P] CTP. The resulting EECs were isolated, and washed. Only short transcripts were produced during the pulse and the labeled tRNAs were removed by the washing protocol. Addition of only NTPs during the 15 min chase generated a

pattern of short and intermediate transcripts that were DRB-insensitive and did not make runoff transcripts. However, addition of K_c cell NE to the chase, allowed some EECs to enter productive elongation and make long DRB-sensitive transcripts. When a HeLa NE that was capable of producing high levels of DRB-sensitive transcription was added to the chase, DRB-sensitive transcripts were also produced.

The HeLa NE used in the other studies in this study was not able to generate an easily detectable level of DRB-sensitive transcripts in the add-back assay, most likely due to the fact that isolated early elongation complexes are not as sensitive to P-TEF as they are when they are formed in the presence of the extract. When α -amanitin was added prior to the chase, elongation by all EECs was inhibited proving that transcripts were produced by RNA polymerase II. When an immobilized HIV-1 DNA template was used to form human EECs both K_c cell and HeLa NE could be added back to form DRB-sensitive transcripts.

3. Effect of Purified *Drosophila* P-TEFb on Transcription

The results presented here show that the HeLa NE used for most studies has limiting amounts of a human P-TEFb -like activity. A number of *Drosophila* basal transcription factors have been shown to be able to replace their human counterparts (Kephart *et al.*, 1994; Wampler and Kadonaga, 1992; Wampler *et al.*, 1990). Purified *Drosophila* P-TEFb was studied to determine if it could function in the human system. Pulse/chase transcription reactions using HeLa NE were carried out. Reactions were chased for either 15 seconds or 1 min and contained either 10 or 40 μ g/ml template.

Addition of *Drosophila* P-TEFb increased the level of DRB-sensitive transcripts present at either time point. The largest increase in DRB-sensitive transcripts was obtained with 40 μ g/ml template. At either template concentration addition of P-TEFb also reduced the amount of short transcripts. The short transcripts also reappeared upon addition of DRB, demonstrating that P-TEFb acts on complexes that are destined to produce short transcripts in the absence of functional P-TEFb.

These results indicate that the CTD kinase activity of P-TEFb is responsible for its effect on transcription and that a large number of phosphorylation events on an EEC by P-TEFb can cause productive elongation. If human P-TEFb is limiting and multiple phosphorylation events
5 are required for the transition into productive elongation, then increasing the number of EECs in a reaction with a constant amount of P-TEFb would cause a decrease in the DRB-sensitive runoff transcripts because the activity of P-TEFb would be spread over a larger number of complexes and the number of EECs that receive the required number of phosphates would be less.

When a template titration from 5 to 40 µg/ml template with a constant amount of HeLa NE was done, the number of EECs increased with the increase in template but the level of DRB-sensitive runoff decreased (FIG. 5). When pure *Drosophila* P-TEFb was added to otherwise identical reactions the level of DRB-sensitive runoff transcripts was dramatically increased. Without additional P-TEFb the level of DRB-sensitive runoff began to decrease above 10 µg/ml, but with additional P-TEFb the level of DRB-sensitive runoff continued to increase at even 40 µg/ml. These results indicate that P-TEFb acts as a CTD kinase that must act multiple times on each EEC.
10
15
20
25

These results demonstrate that two classes of processive human RNA polymerase II elongation complexes are formed after initiation at a promoter *in vitro*. The two classes are differentiated by their elongation rates and sensitivities to DRB. DRB-sensitive transcripts were elongated at a faster rate than the DRB-insensitive transcripts. However, both types of complexes were able to transcribe RNAs up to 4000 nucleotide in length. HeLa extracts made using a protocol similar to that used to produce *Drosophila* K_c cell NE generated a much lower fraction of DRB-insensitive long transcripts. These extracts may have more of the negative factors required or less of the DRB-insensitive factor acting like P-TEFb or may have different levels of general elongation factors. Nevertheless, these results of this example clearly demonstrate the functionality of the human cell system.
20
25

These results indicate that P-TEFb -like activity is limiting in HeLa nuclear extracts for the generation of DRB-sensitive transcripts. Purified *Drosophila* P-TEFb stimulated the production of DRB-sensitive transcripts up to 10-fold in the HeLa transcription system. Therefore, even if other factors are required to generate DRB-sensitive transcripts, these factors
5 are not limiting. High levels of exogenous P-TEFb did not reduce the formation of DRB-insensitive, processive complexes. If there is another factor which causes the transition into DRB-insensitive productive elongation, it must not be in competition with P-TEFb to modify EECs.

10 These results provide further evidence for the action of P-TEFb as a CTD kinase. Multiple phosphorylation events are required to cause the shift from the II_a to the II_b form of the polymerase, and P-TEFb is not stably associated with the elongation complex (Marshall and Price, 1992; Marshall and Price, 1995). Therefore, the distributive action of P-TEFb depends on the concentration of the factor and the concentration of its substrate (EECs). The results presented here substantiate this conclusion. As the template concentration was increased in a background of constant low level of P-TEFb found in HeLa NE, the amount of DRB-sensitive runoff decreased as the number of EECs increased. When P-TEFb levels were elevated, the same template titration now gave more DRB-sensitive runoff at every template concentration and now, the number of DRB-sensitive runoff transcripts did not decrease as the number of EECs increased. These results are consistent with the distributive action of P-TEFb such that at low concentrations of the factor and high template concentrations many of the EECs do not receive the required number of phosphates to cause the transition into productive elongation. When the P-TEFb levels were higher more of the EECs were phosphorylated to the level required for the transition to take place.
15
20

25

EXAMPLE 4
Cloning and Expression of *Drosophila* P-TEFb

A. Peptide Sequences from P-TEFb

5 *Drosophila* P-TEFb was purified as described in Marshall and Price (1995) and Example 1 (Methods, Section 3) from *Drosophila* K_c cell nuclear extract derived from 500 liters of K_c cells (see Example 3, Methods, Section A4 for preparation). About 30 µg of pure P-TEFb was run on a 6-15% gradient SDS-PAGE protein gel. The regions of the gel containing individual subunits were excised and sent to the W.M. Keck Foundation Biotechnology Resource 10 Laboratory (Yale University, New Haven CT). Using standard techniques, each protein was subjected to *in situ* proteolysis and the resulting peptides were separated using reversed phase HPLC. Individual HPLC peaks were analyzed by MALDI mass spectrometry and fractions containing predominately one species were subjected to sequencing. Two or three peptide sequences were obtained from each subunit.

B. Cloning of *Drosophila* P-TEFb Small Subunit

15 Two peptide sequences, MLQQPSGSTPSNV (SEQ ID NO:31) and ADTALNHDFFWTDPMPS (SEQ ID NO:32), were obtained as described above in Section 1 from sequencing the small subunit fragments. Based on these two peptides, two degenerate primers,

20 5'-GGAATTCNATGYTN CARCARCC (SEQ ID NO:9) which encoded the region MLQQP (SEQ ID NO:33) of SEQ ID NO:31 and

5'-AACTGCAGTCCARAARAARTCRTGRTT (SEQ ID NO:10) which encoded the region NHdffWT (SEQ ID NO:34) of SEQ ID NO:32,

25 were designed. The template for the PCR™ reaction was *Drosophila* K_c cell cDNA made through reverse transcription using a 3' RACE™ (Rapid Amplification of cDNA End) kit according to the manufacturer's directions (Gibco, Gaithersburg, MD). A 1.1 kb cDNA fragment was amplified by PCR™ (30 cycles of 0.5 min at 94 °C, 1 min at 55 °C and 1.5 min at 72°C) using Vent™ DNA polymerase. A 0.7 kb fragment from the 3' portion of the PCR™ product was 30 cloned into the Bluescript SK (Stratagene) after digestion with *Pst* I (one internal *Pst* I site in the

PCR™ product and one *Pst I* site designed into the 3' end of the splice site of Bluescript SK (Stratagene).

The 3' region of full length cDNA sequence was obtained using a Gibco 3' RACE™ kit
5 with two specific primers from the 0.7 kb fragment,

5'-TGTCAAGGATCAAACCGGCTGTGAT (SEQ ID NO:11) and
5'-CGAATTCCAAGAAACGCATCGATGC (SEQ ID NO:12).

The 5' region of the full length cDNA was obtained using a Gibco 5' RACE™ kit with three gene specific primers from the 0.7 kb fragment

10 5'-AGACCTGCCAAATCGTGT (SEQ ID NO:13),
5'-AGAAGGTGGATCTGTAACCATTCTGT (SEQ ID NO:14) and
5'-GGAATTCAAGATCTCGATCAGATTCA (SEQ ID NO:15).

15 The coding sequence of the small subunit was cloned by reverse transcription PCR™ (RT-PCR). First, the cDNA of small subunit was generated in a reverse transcription using a primer 5'-TTACTACTCGAGCTACCAAACCCGGTC (SEQ ID NO:16) and *Drosophila* embryonic mRNA as the template. Second, the coding sequence was produced in a 30-cycle PCR™ reaction using VENT DNA polymerase, two primers

20 5'-TAAGCAAGCTTCTATGGCGCACATGTCC (SEQ ID NO:17) and
5'-TTACTACTCGAGCTACCAAACCCGGTC (SEQ ID NO:18)

and the cDNA as the template. The PCR cycles were performed for 0.5 min at 94°C, 1 min at 55°C and 3.5 min at 72°C according to the manufacturer's directions (NEN, Boston, MA). Finally, the coding sequence was digested with *Hind III* and *Xho I* and cloned into pET-21a vector.

25

C. Cloning of *Drosophila* P-TEFb Large Subunit:

Three peptide sequences SPEWPDI (SEQ ID NO:35), WYFSNDQLANSPSR (SEQ ID NO:36) and TVHGMPPFEQQLPY (SEQ ID NO:37) of the large subunit that were obtained as

described above in Example 4, Section 1. Based on these amino acid sequenced, three degenerate primers,

5'-GGAATTCTGGTAYTTYWSNAAYGA (SEQ ID NO:19) which encoded the region WYFSND (SEQ ID NO:38) of SEQ ID NO:36,

5 5'-CGGGATCCTGYTCRAANGNGGCAT (SEQ ID NO:20) which encoded the region MPPFEQ (SEQ ID NO:39) of SEQ ID NO:37, and

5'-CGGGATCCAANGNGGCATNCCRT (SEQ ID NO:21) which encoded the region HGMPPF (SEQ ID NO:40) of SEQ ID NO:37 were designed.

10 A 1.6 kb cDNA fragment of the large subunit was cloned through nested PCR™ reactions. First, a 35-cycle PCR™ reaction was performed using Taq DNA polymerase, the degenerate primers, SEQ ID NO:19 and SEQ ID NO:20, and the *Drosophila* embryonic cDNA obtained from Clontech (Palo Alto, CA) as the template. The PCR™ cycles were performed for 1 min at 94°C, 1 min at 55°C and 3 min at 72°C. Second, a 1.6 kb cDNA fragment was amplified in another 35-cycle PCR™ reaction using the total PCR™ products from the first amplification as the template, the primers of SEQ ID NO:19 and SEQ ID NO:21 and the same reaction conditions.

15 20 The 1.6 kb fragment of the second round of amplification was digested with *EcoRI* and *BamHI* to yield three fragments of 0.9 kb, 0.6 kb and 0.1 kb. The 0.9 kb and 0.6 kb fragments were cloned into Bluescript SK (Stratagene), yielding a 1.5 kb insert, and sequenced using fluorescent automated sequencing at the DNA Facility at the University of Iowa, following manufacturer's recommended protocols.

25 The 3' region of the full length cDNA sequence was obtained using the 3' RACE kit (Gibco) according to the manufacturer's directions. Three nondegenerate, specific primers,

5'-ATCACGACACCACCAAGAGCTGTTA (SEQ ID NO:22),

5'-CGAATTCAAGATCGTGAACGGGA (SEQ ID NO:23) and

5'-CGAATTCAAGCGCTAGCAATG (SEQ ID NO:24),

were designed based on the 1.5 kb cDNA sequence obtained. The 5' region of cDNA sequence was obtained using a 5' RACE™ kit (Gibco) according to the manufacturer's directions. Three gene specific primers,

5'-GAAAGGCGTAGAACCGA (SEQ ID NO:25),

5 5'-GCTGACCCATTCCTGTATCAGATAG (SEQ ID NO:26) and

5'-GGAATTCTCTGCTGGCGAAT (SEQ ID NO:27),

were designed based on the 1.5 kb sequence and used with the 5' RACE kit (Gibco).

The entire coding sequence of the large subunit was cloned by reverse transcription PCR™ (RT-PCR). First, the cDNA of large subunit was generated in a reverse transcription reaction using the primer, 5'-GGGAATTCTGAGGTTCTATACATAT (SEQ ID NO:28) and *Drosophila* embryonic mRNA as the template. Second, a 4 kb cDNA fragment containing the coding sequence was produced in a 35-cycle PCR™ reaction using Expand™ polymerase (Boehringer Mannheim, Indianapolis, IN), the two primers

15 5'-CTGTGTGAATGGAATCTGTGATGTG (SEQ ID NO:29) and

5'-GGGAATTCTGAGGTTCTATACATAT (SEQ ID NO:28)

and the *Drosophila* cDNA as the template. The PCR™ cycles were performed according to the following protocol, supplied by the manufacturer:

20 Amplification reactions were denatured for 2 min at 94°C; 10 cycles, 10 sec. at 94°C, 30 sec at 58°C, 3 min at 68°C; 5 cycles, 10 sec at 94°C, 30 sec at 58°C, 4.5 min at 68°C; 5 cycles, 10 sec at 94°C, 30 sec at 58°C, 6 min at 68°C; 5 cycles, 10 sec at 94°C, 30 sec at 58°C, 7.5 min at 68°C; 5 cycles, 10 sec at 94°C, 30 sec at 58°C, 9 min at 68°C; 1 cycle, 7 min at 68°C.

The resulting amplification product was reamplified in another PCR™ reaction using Vent™ 25 DNA polymerase, the reamplified 4 kb cDNA fragment and the primers

5'-TATCCCGGGTCATATGAGTCTCCTAGCC (SEQ ID NO:30) and

5'-GGGAATTCTGAGGTTCTATACATAT (SEQ ID NO:28).

Finally, the coding sequence was digested with *Sma I* and *EcoR I* and cloned into pET-21a vector.

D. Characterization of P-TEFb

cDNA fragments of P-TEFb large and small subunits were amplified from the total *Drosophila* embryonic cDNA by using the degenerate primers derived from the peptide sequences, see Sections B and C above. Based on the cDNA fragments, the 5' end and 3' end sequences of P-TEFb cDNAs were obtained from *Drosophila* K_c cell mRNAs using the RACE technique. The full length small subunit cDNA sequence (SEQ ID NO:1) was 1.45 kb containing an open reading frame of 1.21 kb, beginning at position 115 of SEQ ID NO:1. The large subunit cDNA (SEQ ID NO:3) was 4.33 kb with a coding region of 1.10 kb, beginning at position 716 of SEQ ID NO:3. The encoded small subunit was found to be a cyclin-dependent kinase (CDK) of 46.8 kDa (SEQ ID NO:2), while the large subunit was a cyclin of 121.1 kDa (SEQ ID NO:4). All data derived from the sequences are consistent with the identified properties of P-TEFb.

The amino acid sequence, SEQ ID NO:2, identified the small subunit of *Drosophila* P-TEFb as a member of the Cdc2-like cyclin dependent kinase family with over 40% identity to *S. pombe* Cdc2. Thus P-TEFb is a cyclin-dependent kinase (CDK) because the small subunit has all the conserved subdomains found in CDKs and the large subunit has the conserved cyclin box domain. The kinase activity of a CDK is tightly regulated by four conserved mechanisms. A CDK can be activated by the binding of a cyclin subunit and the phosphorylation of a conserved threonine residue at the T-loop in the catalytic subunit. A CDK-cyclin complex can be either inhibited by the phosphorylation of a threonine residue and a tyrosine residue at the ATP-binding site in the catalytic domain, or inactivated by the binding of a family of small proteins termed CKIs (Morgan, 1995).

25

Whereas the peptide sequences of catalytic subunits are well conserved in the CDK family, the sequences of cyclins are rather divergent. Almost all of the cyclin subunits contain a diverse sequence of about 100 residues called a cyclin box, which is predicted to have a

conserved helix-rich secondary structure. Helical fold prediction for the cyclin box is described in Proteins 24, 1-17.

The full length cDNA of the *Drosophila* large subunit encoded a 1097 amino acid protein (SEQ ID NO:4). As expected the protein contained a canonical cyclin box presumably to allow binding to the small cdk subunit. The carboxyl-terminal two thirds of the protein did not reveal any specific protein motifs, however, the middle third is highly charged and likely to be somewhat unstructured, while the carboxyl-terminal third has high levels of potential helical regions suggesting that it might fold into a specific domain. Attempts to express the cDNA in *E. coli*, even with the small subunit, only gave insoluble truncated proteins. However expression of both subunits in a baculovirus expression system gave rise to recombinant proteins with identical mobility to authentic P-TEFb purified from *Drosophila* K_c cells on a silver stained SDS PAGE gel. Purification of the recombinant protein was aided by the addition of a HIS-tag to the carboxyl-terminus of the small subunit. Both subunits were quantitatively recovered on a nickel column indicating that the subunit interactions were strong. This protein had levels of DRB-sensitive CTD kinase activity indistinguishable from authentic P-TEFb and was able to functionally replace authentic P-TEFb during transcription.

Northern blotting was used to further confirm the presence of the P-TEFb mRNAs in *Drosophila* K_c cells, embryos and female adults. A probe made from the full length of the small subunit cDNA recognizes a mRNA band around 1.7 kb. Three probes that were generated from the 3', internal, 5' cDNA sequences of the large subunit cDNA recognize an mRNA band approximately 5.3 kb in *Drosophila* adult females. Final wash stringency was 0.1× SSC at 68°C.

25 E. Expression of P-TEFb in *E. coli*

The full length P-TEFb cDNAs were amplified from embryonic mRNA by reverse transcription PCR™ (RT-PCR™) and cloned into a pET21a expression vector in *E. coli* as described above in Sections B and C. Constructs were made to express (1) the small subunit, (2) a GST-small subunit fusion protein, (3) the large subunit, and (4) the GST-large subunit fusion

protein. Two other constructs were used to co-express the small subunit and the large subunit, or the small subunit and the GST-large subunit fusion protein.

Six constructs were made to express P-TEFb in *E. coli* using a T7 polymerase-dependent expression system. First, a GST coding sequence from pEG(KT) was used to replace the *Nde I/Sal I* fragment in pET21a (Novagen) to construct a GST-expression plasmid (pET21a-GST). Second, the large subunit coding sequence was digested with *SmaI* and *EcoRI* and cloned into pET21a-GST vector to construct a plasmid (pET21a-GST-BL) for expression of the GST-large subunit fusion protein. The small subunit coding sequence was digested at the designed *Hind III* site and *Xho I* sites and cloned into pET21a-GST vector to construct a plasmid (pET21a-GST-BS) for expression of the GST-small subunit fusion protein. Third, the GST coding sequence was removed from the pET21a-GST-BL and the pET21a-GST-BS by *Nde I* digestion and the recombinant vectors were religated to construct a plasmid (pET21a-BL) for expression of the large subunit and the other plasmid (pET21a-BS) for expression of the small subunit.

The final constructs were prepared by obtaining the coding region of the GST-large subunit fusion protein via the *Xba I* digestion of pET21a-GST-BL, and inserted into pET21a-BS to construct a plasmid (pET21a-GST-BL-BS) for the co-expression of the GST-large subunit fusion protein and the small subunit. The coding region of the large subunit was obtained by the *Xba I* digestion of pET21a-BL, and inserted into pET21a-BS to construct a plasmid (pET21a-BL-BS) for the co-expression of the large subunit and the small subunit.

All constructs gave rise to appropriately sized protein products when transformed into the DE3 host and induced by adding isopropylthio-beta-galactoside (IPTG); P-TEFb small subunit (43 kDa); GST-P-TEFb small subunit (73 kDa); P-TEFb large subunit (124 kDa); GST- P-TEFb large subunit (154 kDa). All proteins were insoluble when expressed alone or in combinations with both large and small subunits.

Two constructs were made to express P-TEFb in the insect cells by using a Baculovirus expression system. First, the coding region of the small subunit was obtained by the *Hind III* and *Xho I* double digestion of pET21a-GST-BS, and inserted into pBAC4X-1 (from Novagen) to construct a plasmid (pBAC4X-1-BS) for the expression of the small subunit. Second, the coding region of the large subunit was obtained by the *Xma I* and *EcoR I* double digestion of pET21a-GST-BL, and inserted into pBAC4X-1-BS to construct a plasmid (pBAC4X-1-BL-BS) for the co-expression of the large subunit and the small subunit.

The protocol for expression in Sf9 cells is as follows:

1. Linearized BaculoGold DNA and transfer vector:

a. DNA templates:

Linearized BaculoGold DNA (2.5 ug/25 ul)

Transfer vector containing large and small subunits of *Drosophila* P-TEFb
pBAC4X-1-BL-BS (1 ug/ul)

b. Buffers (see Example 3, Section A.2):

Buffer A (Grace's Medium with 10% Fetal Calf Serum)

Buffer B (25 mM Hepes pH 7.1, 125 mM CaCl₂, 140 mM NaCl)

2. Co-transfection:

Sf9 cells are cultured in TNM-FH medium (in suspension) to log phase. 2 million cells are placed in an T-25 flask and the cells are allowed to attach to the bottom of the flask for 30 min. The medium is removed from the flasks and 1 ml of Buffer A is added to each flask such that the buffer covers all cells. 4 ug transfer vector pBAC4X-1-BL-BS and 0.5 ug BaculoGold DNA are mixed together in a sterile tube for 5 min. 1 ml of Buffer B is added to each tube containing the DNA mixture. All of the DNA solution (now in Buffer B) is pipetted, drop-by-drop, onto the cells such that the buffer covers all cells. Minor precipitation should be visible in the flask. Cells are left undisturbed at 27°C for 4 h. After 4 h the incubation medium is changed.

The medium is removed from the transfection and control flasks. 3 ml TNM-FH + 10% FCS is added to each flask and cells are incubate at 27°C for ~4-5 days.

5

EXAMPLE 5 Human P-TEFb

A. The Small, Kinase Subunit

The human homolog of the small subunit of *Drosophila* P-TEFb was first identified by comparing the protein sequence derived from direct protein sequencing of the small subunit of *Drosophila* P-TEFb with genetic databases accessible via a BLAST genetic analysis search from the National Institutes of Health (NIH). Using the complete sequence of the *Drosophila* subunit obtained from translation of a full length cDNA, the inventor identified the existence of a human cDNA homologous to the *Drosophila* P-TEFb small subunit protein. The search of the protein database revealed a human protein, PITALRE, SEQ ID NO:6 (Grana et al., 1994), that exhibits about 72% identity and about 83% similarity to the *Drosophila* protein, SEQ ID NO:2. The high level of sequence similarity indicated that PITALRE is a potential homologue of the small subunit of *Drosophila* P-TEFb and, therefore, may be a component of human P-TEFb. Two kinases from *S. cerevisiae*, SGV1 (Irie et al., 1991) and CTK1 (Sterner et al., 1995), each share 43% identity with PITALRE, SEQ ID NO:6, and the small subunit of *Drosophila* P-TEFb, SEQ ID NO:2. Although sequence similarity does not allow the prediction of a potential yeast homologue, CTK1 has recently been demonstrated to increase the elongation efficiency of RNA polymerase II (Lee and Greenleaf, 1997).

The human protein identified was first cloned using PCR™ with degenerate oligonucleotides derived from cell division cycle 2 (CDC2) family sequences (Grana et al., 1994). This protein was called "PITALRE" because of the presence of those amino acids in a characteristic location in the kinase subunit. The initial characterization of "PITALRE" included a sequence comparison with other kinases, a northern blot showing ubiquitous expression, an immunoprecipitation study suggesting the association of other proteins with "PITALRE", and a western blot suggesting that the protein was localized in the nucleus. The highest expression of

the protein was in the liver and placenta which caused the authors to speculate that "PITALRE" was involved in specialized functions in certain cell types (Grana *et al.*, 1994). Although Grana *et al.* (1994) investigated the function of the protein "PITALRE" and its interaction with the tumor suppresser gene product pRB, they did not suggest that it may interact with RNA polymerase II. A later paper mapped the chromosomal location of "PITALRE" to a region found to be involved in tumors and breast cancer (Bullrich *et al.*, 1995).

B. The Large, Cyclin-like Subunit

The inventor realized that human P-TEFb, like the *Drosophila* protein, is likely to comprise at least two subunits, one a cyclin-dependent kinase (CDK) and the other containing a cyclin box domain. As the amino terminal portion of the large subunit has a cyclin box, the inventor contemplated that this portion of the large subunit interacts with the small kinase subunit. The carboxyl terminal portion of the large subunit is larger than CTDs previously found attached to cyclin boxes, leading the inventor to envision that this portion of the large subunit interacts with other factors, for example the HIV viral protein Tat.

1. Cloning of Human P-TEFb Large Subunits:

Considerable difficulty was encountered in the initial cloning of the cyclin subunits of human P-TEFb, although the ultimate success of the inventor means that this can now be routinely achieved. Initially, searches of protein sequence databases revealed no homologous proteins of the *Drosophila* P-TEFb large subunit. Only by searching the EST database, with the BLAST genetic analysis (National Institutes of Health), for homologues of the *Drosophila* P-TEFb large subunit were three, relatively short, human EST sequences, zr91f10.s1, yd48c03.r1 and nc70h05.r1, found that the inventor realized may be part of the human P-TEFb large subunit.

25

Without having the *Drosophila* large subunit sequence it would have been impossible to identify any human homologues. The EST database contains very short sequences with numerous errors. For example, high quality sequence for zr91f10.s1 stops at 214 bases; for yd48c03.r1 at 204 bases; and for nc70h05.r1 at 339 bases. Further, the EST clones often

contained 5' UTR, 3' UTR or unspliced sequence that could not be compared to the *Drosophila* sequence. Even when the EST sequences were found to encode protein sequence similar to the *Drosophila* sequence, errors in the sequence caused the reading frame to be lost. Without the *Drosophila* sequence it would have been impossible to know whether the sequence was important or not. Each of the EST sequences does represent a human cDNA sequence, but no function was hypothesized for any of the EST sequences. Further clone distribution information for all three EST sequences can be found through either the National Center for Biotechnology Information at the National Institutes of Health or through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html.

10

Based on the sequence from zr91f10.s1, three primers,

5'-TTCCCACCAATGCTTCC-3' SEQ ID NO:51, 5'-CCATCAGTTGATAACAGGGATCT-3' SEQ ID NO:52, and 5'-GGAATTCAAGAAGGTTGTAAGATGC-3' SEQ ID NO:53 were designed and used to obtain the 5' region of a total cDNA sequence from human brain poly A+ RNA using a 5' RACE kit (Gibco). The 3' region of a total sequence was obtained by using three primers, 5'-ACACACAGATGTGGTGAAATGTACCCA-3' SEQ ID NO:54, 5'-GCATCTTACAACCTTCTG-3' SEQ ID NO:55, and 5'-GGAATTCATGGAAAGCATTGGTGGGAAT-3' SEQ ID NO:56, a brain Marathon-ready cDNA and a Marathon cDNA Amplification Kit (Clontech). The total cDNA sequence obtained was 4528 bp and termed HBL1, SEQ ID NO:43.

The total cDNA sequence, HBL1, (SEQ ID NO:43) was amplified by RT-PCR using primers 5'-CCTCCACTACTGGTTGCCTGG-3' SEQ ID NO:57, 5'-GGACTAGTATAATATGGCGTCGGGCCGTG SEQ ID NO:58, and 5'-GGAGATCTTACATGTTCATTCCCTGGG SEQ ID NO:59 and Expand polymerase under the following conditions:

step 1	94°C	2 min
step 2	94°C	10 sec
step 3	65°C	30 sec
step 4	68°C	3 min

step 5	Go to step 2 for 9 times	
step 6	94°C	10 sec
step 7	65°C	30 sec
step 8	68°C	4.5 min
step 9	Go to step 6 for 4 times	
step 10	94°C	10 sec
step 11	65°C	30 sec
step 12	68°C	6 min
step 13	Go to step 10 for 4 times	
step 14	94°C	10 sec
step 15	65°C	30 sec
step 16	68°C	7.5 min
step 17	Go to step 14 for 4 times	
step 18	94°C	10 sec
step 19	65°C	30 sec
step 20	68°C	9 min
step 21	Go to step 18 for 4 times	
step 22	68°C	7 min
step 23	stop	

Two coding sequences were amplified: HBL1-1 (2091 bp), SEQ ID NO:44, and HBL1-2 (2190 bp), SEQ ID NO:46. They were cloned in a plasmid pBAC 4X-1-HBS, which contains human P-TEFb small subunit sequence. These two sequences were also amplified by using HeLa RNA.

Based on the EST sequence of nc70h05.r1, the three primers

5'-GGAGACAAGTATGTGCTACCTTGATGACA-3' SEQ ID NO:60,

5'-GGAATTGGGGCTGCTCCTCCACTTAG-3' SEQ ID NO:61, and

5'-GGAATTGCGCTGGAGGCCACAGAA-3' SEQ ID NO:62 were used to obtain the 5' region of a total cDNA sequence from human bone marrow Marathon-ready cDNA (Clontech) by using a Marathon cDNA Amplification Kit (Clontech). Using the same cDNA and RACE kit, the 3' region was obtained by using the primers 5'-GTGTCACTGAAAGAATACCG-3' SEQ ID NO:63 and 5'-GGAATTCAAGGTGGAGATAAGCTGC-3' SEQ ID NO:64, which were based on the EST sequence of yd48c03.r1. A total cDNA sequence was obtained and designated HBL3 (SEQ ID NO:48). According to the length of the PCR products, the whole HBL3 was 2.8 kb.

The first 2.36 kb of the PCR product was sequenced. The rest of the sequence was 3' UTR and was not sequenced.

5 The total cDNA sequence, HBL3, (SEQ ID NO:48) was then amplified by RT-PCR using the primers 5'-GCTCTAGATAAAATATGGAGGGAGAGAGGAA-3' SEQ ID NO:65,
5'-GGAATTCTTACTTAGGAAGGGGTGGAAGTG-3' SEQ ID NO:66, and
5'-GGAATTCTTACTTAGGAAGGGGTGGAAGTGGTGGAGGAGGTT-3' SEQ ID NO:67 from human HeLa cell mRNA. Conditions for amplification of the large subunit using eLONGase (Life technologies) were denaturation for 30 sec at 94°C; then 35 cycles, 20 sec at
10 94°C, 30 sec at 55°C, 2.2 min at 68°C; then 5 min at 68°C for each reaction.

A single coding sequence for HBL3 (2181 bp), SEQ ID NO:49, was amplified and cloned into a plasmid pBAC 4X-1-HBS as described previously.

15 **2. Analysis of Human P-TEFb Large Subunit cDNAs and Proteins:**

20 The coding sequence HBL1-1 (SEQ ID NO:44) encodes one complete protein (SEQ ID NO:45). Comparison of the cDNA sequences of HBL1 (SEQ ID NO:43) and HBL1-1 shows that the entire coding sequence HBL1-1 is contained within HBL1. The beginning of the start codon (ATG) of the encoded protein corresponds to position 1 of HBL1-1 and position 46 of HBL1, respectively. Thus, the total cDNA sequence of HBL1 includes 45 bp upstream of the start codon and 2392 bp downstream of the stop codon (TAA). Interestingly, HBL1-2 (SEQ ID NO:46) contains an 99 bp intron which is not present in either HBL1 (SEQ ID NO:43) or HBL1-1 (SEQ ID NO:44) but is in-frame and can be translated. This 99 bp intron begins at position 1927 and ends at position 2025 of coding sequence HBL1-2 (SEQ ID NO:46). If the 99 bp
25 intron of HBL1-2 is excluded then HBL1-1 and HBL1-2 are identical.

30 The coding sequence of HBL3 (SEQ ID NO:49) encodes a different protein which is completely contained with the total cDNA sequence of HBL3 (SEQ ID NO:48). The beginning of the start codon (ATG) of the encoded protein corresponds to position 1 of the coding sequence of HBL3 (SEQ ID NO:49) and position 45 of the total cDNA sequence of HBL3 (SEQ ID

NO:48), respectively. Thus, the total cDNA sequence of HBL3 includes 44 bp upstream of the start codon and 135 bp downstream of the stop codon (TAA).

Comparison of the total cDNA sequences, HBL1 (SEQ ID NO:43) and HBL3 (SEQ ID NO:48), revealed that the EST sequence zr91f10.s1 was homologous to a small region of HBL1 and a perfect match to the compliment of positions 603-665 of SEQ ID NO:43. It is noteworthy that only 50 bp of the *Drosophila* large subunit sequence is homologous to this same EST sequence zr91f10.s1. Whereas, the EST sequences nc70h05.r1 and yd48c03.r1 were homologous to HBL3 and near perfect matches to positions 210-417 and 750-884, respectively, of SEQ ID NO:48.

Comparison of the encoded proteins showed that the proteins encoded by HBL1-1 and HBL1-2, SEQ ID NO:45 and SEQ ID NO:47, respectively, are identical except for a 33 bp contiguous stretch which is encoded by the intron in HBL1-2. Comparison of the protein encoded by HBL1-1, SEQ ID NO:45, and the protein encoded by HBL3, SEQ ID NO:50, shows that the two human proteins share an overall absolute identity of 54% and an overall relative similarity of 70%. Within the region defined as the cyclin box, positions 1-252 of HBL1-1 and positions 1-253 of HBL3, the two proteins share an identity of about 81% to each other. Downstream of the cyclin box, the two proteins, SEQ ID NO:45 and SEQ ID NO:50, are about 46% identical to each other.

When compared to the *Drosophila* large subunit protein, SEQ ID NO:4, both human proteins have an identity of about 65% to the *Drosophila* cyclin domain box (positions 1-280 of SEQ ID NO:4). The two human proteins were 81% similar to each other in this same region. However, downstream of the cyclin box, the two human proteins are only about 25% identical to the *Drosophila* protein. Overall, the *Drosophila* large subunit is about 42% identical to the protein encoded by HBL1-1 (SEQ ID NO:45) and 34% identical the protein encoded by HBL3 (SEQ ID NO:50).

P-TEFb
1996-08-20

As was seen for the *Drosophila* large subunit the amino-terminal domain contained a cyclin box that was 65% identical to the *Drosophila* protein. The two human proteins were 81% similar in this region. Both potential human proteins lacked the highly charged, unstructured central region, but had a slight similarity (25% identity) to the carboxyl-terminal domain.

5 Western blotting with antibodies to each of the two proteins that were expressed in *E. coli* are used to determine the size of both encoded proteins. The two cyclin subunits were co-expressed with a HIS-tagged human kinase subunit (PITALRE) in a baculovirus system. As was found for *Drosophila* P-TEFb the large subunit was quantitatively recovered after nickel column chromatography indicating that subunit interactions are strong. In contrast to PITALRE alone,

10 both kinase/cyclin pairs (PITALRE/HBL1-1 and PITALRE/HBL3) gave rise to very strong DRB sensitive phosphorylation of the CTD of RNA polymerase II. This clearly indicates that human P-TEFb is a cyclin dependent kinase, since the activity of PITALRE was stimulated approximately 100 fold by the cyclin subunit. The two subunit protein obtained after expressing HBL1-1 large subunit with PITALRE was able to function in transcription when added back to a

15 *Drosophila* nuclear extract depleted of P-TEFb using anti large subunit antibodies. Although it is not clear if HBL1-1 encodes the proper length protein, HBL3 encodes a protein with identical mobility to the 87 kDa protein found in PITALRE immunoprecipitates. This is confirmed by using antibodies for the human proteins and western blot analysis. It is likely that antibodies against the protein encoded by HBL3 react with the 87 kDa band seen in PITALRE immunoprecipitates.

20

3. Expression of Human P-TEFb:

The coding regions for HBL1-1, HBL1-2 and HBL3 were individually expressed in a Baculovirus expression system (PharmMingen, San Diego CA) along with the HIS-tagged small subunit "PITALRE". A nickel column was used to purify the "PITALRE" from the cytoplasm of infected Sf9 cells. The "PITALRE" was found to associate with each respective putative large subunit protein, and thus indicates that the cyclin domain of each protein can interact with the kinase subunit.

Using the CTD Kinase assay, as previously described in Example 2, Section A.7, both HBL1-1 (SEQ ID NO:45), HBL2(SEQ ID NO:47) and HBL3 (SEQ ID NO:50) were substituted for the large subunit in the *Drosophila* transcription system. The activity of each recombinant protein is shown in Table 1. Alone "PITALRE" has little kinase activity and no detectable activity in transcription. Thus, these data demonstrate that the proteins encoded by HBL1-1, HBL1-2 and HBL3 can functionally act as the P-TEFb large subunit in *Drosophila* and likely act in a homologous manner in humans.

Table 1. Recombinant proteins expressed *in vitro* using a kinase assay and a HIS-tagged small subunit.

Subunits	Complex formed?	Kinase activity	Transcription activity
DBS	-	very little	none
HBS	-	very little	none
DBS+DBL	yes	high	high
HBS+DBS	no	-	-
HBS+HBL1	yes	high	high
HBS+HBL2	yes	ND	ND
HBS+HBL3	yes	high	ND

D is *Drosophila*, H is human, BS is P-TEFb small subunit, BL is P-TEFb large subunit, ND is not determined. If the large subunit was recovered along with the HIS-tagged small subunit from the nickel column it was assumed that a complex was formed. The kinase assay used RNA polymerase II as a substrate and transcription activity was determined by adding the recombinant protein into a *Drosophila* extract depleted of P-TEFb using antibodies against the large subunit.

4. Tissue Distribution of Expression of Human P-TEFb Subunits.

The expression of mRNAs encoding PITALRE and the two human cyclin subunit clones was examined across a wide variety of human tissues using northern blot analyses. All three were ubiquitously expressed. This is consistent with a general requirement for P-TEFb in all tissues. mRNAs encoding both cyclin subunits were present in all tissues suggesting that a tissue specific role for one of the subunits is not likely.

5. Reconstitution of Tat Transactivation with Recombinant Human P-TEFb.

To confirm that the recombinant human P-TEFb comprised of PITALRE and cyclin HBL1-1 could reconstitute Tat transactivation, an add back study was performed on HeLa

extracts depleted of PITALRE and, therefore, depleted of P-TEFb. Conditions were as described in Example 6. FIG. 6 shows that Tat was able to stimulate the DRB sensitive runoff transcript from the HIV-LTR.

5 **6. Generation of an Inducible PITALRE Kinase Knockout HeLa Cell Line.**

A tetracycline inducible HIS-tagged PITALRE kinase knockout (D167N) expression vector was constructed and stably transfected Tet-on HeLa cells using co-transfection with a hygromycin expression plasmid. Cell lines were selected in the presence of hygromycin, but in the absence of doxycyclin. Several lines were obtained that exhibited doxycyclin induced expression of a protein with slightly lower mobility than PITALRE that reacted with anti-PITALRE antibodies. A time course indicated that after 30 hrs of induction maximal levels of the inactivated PITALRE were produced. The level of the induced PITALRE mutant was equal to the level of endogenous PITALRE as analyzed by SDS PAGE followed by immunoblotting with anti-PITALRE antibodies. A cell line that expressed a great excess of transfected PITALRE was not obtained, suggesting that the levels of PITALRE may be controlled by post-transcriptional mechanisms.

10 **7. Optional Purification of Human P-TEFb:**

Purification of human P-TEFb is simplified by the use of antibodies that detect the small subunit of P-TEFb. The inventor has generated rabbit polyclonal antibodies that react with the small subunit of P-TEFb. These antibodies were raised against a recombinant human P-TEFb small subunit. A cDNA encoding the human P-TEFb small subunit was amplified from human cDNA using PCR™ primers designed from the human cDNA sequence found in the database. The cDNA was cloned in a pET expression plasmid and the resulting plasmid was expressed in 20 DE3 cells. The recombinant protein was purified under denaturing conditions and subsequently used to inoculate rabbits. The antibodies were generated using standard techniques (Pocono 25 Rabbit Farm).

On western blots the antibodies strongly recognize the bacterially produced human 30 protein and a 42 kDa protein in HeLa nuclear extracts. The inventor has analyzed HeLa extracts

that have functionally different levels of P-TEFb activity and has found that the differences correlate with the level of the 42 kDa protein present in the extracts. Immunoprecipitation and immunodepletion studies are useful in showing that functional human P-TEFb (transcriptional activity and CTD kinase activity) is recognized by the antibodies.

5

Purification of human P-TEFb from HeLa cell nuclear extract is carried out by using standard methods, essentially as described in Examples 1 and 2 and Marshall and Price (1995). P-TEFb is assayed after each step of the purification using a western blot. Evidence already shows that the purification of human P-TEFb is achievable using methods similar to those used
10 for the purification of the *Drosophila* protein. Chromatography on phosphocellulose was carried out and it was found that P-TEFb eluted in the high salt fractions. The inventor's results are consistent with the existence of more than one chromatographic form of P-TEFb. Further purification is achievable using chromatography on phenyl-sepharose, Mono Q™, Mono S™ and hydroxylapitite.

15

Although the data is consistent with the existence of multiple forms of the human large subunit, the form or forms isolated in the 42 kDa HeLa extracts correlate with P-TEFb activity.

20
25

EXAMPLE 6 **P-TEFb is the Tat-Associated Kinase**

As P-TEFb is sensitive to DRB which is canonically a kinase inhibitor, it was clear that P-TEFb is a kinase (Marshall and Price, 1995; Example 1). Although likely targets for a transcription factor kinase that acts early during elongation would be RNA polymerase II or a
25 basal initiation factor such as TFIIF, there is no obvious similarity between kinases known to be involved in transcription regulation and the subunit composition of P-TEFb (Marshall and Price, 1995). Therefore it is a surprising discovery that P-TEFb is essential for the HIV viral protein Tat to activate elongation of the viral RNA genome. This finding is particularly surprising in that it was earlier reported that the TAK protein appear to be most similar to factor 2 or P-TEFa,
30 not P-TEFb (Marshall and Price, 1995).

Tat is a viral protein that acts as an activator of transcription of the HIV viral transcription unit. It binds to short nascent HIV transcripts and causes RNA polymerase II to synthesize mRNA sized transcripts. It was hypothesized by several investigators that Tat might enhance the action of RNA polymerase II CTD kinases and that this might have a positive effect on the elongation potential of the polymerase. An interaction between Tat and a CTD kinase has been demonstrated using immobilized Tat and HeLa nuclear extract (Yang, *et al.*, 1996). The human immunodeficiency virus Tat proteins specifically associate with TAK *in vivo* and require the CTD of RNA polymerase II for function. This TAK protein is sensitive to DRB, but has not been purified or otherwise identified and has not been shown to be required for Tat transactivation.

Similar studies to the immediately preceding were performed using the 48 amino acid transactivation domain of Tat (Tat 48 Δ) fused to GST (GST-Tat) and coupled to glutathione beads. HeLa extracts were incubated with the beads and then the beads were extensively washed. The proteins that were bound were eluted with SDS and heat and analyzed on a gradient protein gel. Silver staining indicated that less than 1% of the HeLa proteins bound, but a significant portion of the human P-TEFb was bound as detected by western blotting using antibodies against the small subunit of human P-TEFb.

A. Methods

1. Tat Proteins:

GST-Tat 1 48 Δ and GST-Tat 1 48 Δ P181S in the pGEX2T vector (Pharmacia) were obtained from the AIDS Research and Reference Reagent Program (NIH) and expressed in *E. coli*. 500 ml cultures were induced and the crude lysates obtained after using the French press were analyzed on silver stained SDSPAGE gels. Roughly equal levels of the 35 kDa proteins were expressed (1 mg/ml of lysate).

2. Binding Assays:

25 µl of glutathione beads were washed 4 times with 200 µl EBC (50 mM Tris, pH 8, 150 mM NaCl, 0.5% NP-40) and then incubated with 100 µl of an *E. coli* lysate containing GST expressed from a pET21 vector (see Example 4). To pre-clear the extract the GST beads were
5 washed 4 times with EBC and then incubated for 3 h. at 4°C with 100 µl of HeLa nuclear extract that had been spun at 15,000 rpm for 5 min. 5 µl of glutathione beads (10 µl of 50% slurry) was transferred to a yellow pipet tip plugged with glass wool and then washed with 400 µl EBC + 5 mM DTT. 5 µl of *E. coli* lysate containing either GST-Tat 1 48Δ or GST-Tat 1 48Δ P181S (controls were done with no bound protein and with GST alone) and 50 µl EBC + 5 mM DTT
10 was allowed to slowly flow over the beads 4 times. The beads were then washed with 200 µl EBC + 5 mM DTT + 0.075% SDS and then 100 µl EBC + 5 mM DTT. 15 µl pre-cleared HeLa nuclear extract was passed over the tip 6 times. The beads were then washed with 900 µl EBC + 5 mM DTT + 0.03% SDS. 20 µl of protein gel loading buffer was added and the tip boiled for 4 min. 10% of the eluted proteins were analyzed on a 6-15% polyacrylamide SDS protein gel by silver staining and 90% of the sample was run on a similar gel and blotted to nitrocellulose. The blot was probed with a 1:1000 dilution of anti-human P-TEFb antisera followed by 1:20,000 dilution of the secondary antibody. Reacting proteins were detected using a ECL detection kit.
15

3. *In vitro* Transcription Conditions

One preferred method to examine the functional interaction between Tat and P-TEFb is to first develop a P-TEFb-dependent human *in vitro* transcription system and then examine the function of Tat with and without added purified human P-TEFb. The P-TEFb-dependent system can be obtained in any of three ways.

25 First since human P-TEFb has similar chromatographic properties to *Drosophila* P-TEFb, it is possible to remove the factor by passing extracts through phosphocellulose at 0.4 M HGKEDP as was done for the *Drosophila* extracts (Marshall and Price, 1995; Example 1). The HeLa-FT (flowthrough) fraction contains much less P-TEFb and is dependent on added P-TEFb to obtain DRB-sensitive transcription.

The second method uses is based on antibody depletion. Purified IgGs from the rabbit anti-P-TEFb serum are covalently coupled to Affigel (BioRad). The IgG beads are washed thoroughly with PBS and then incubated at room temperature with HeLa nuclear extract. The 5 beads are then removed by centrifugation or filtration and the resulting extract is examined for a reduction in DRB-sensitive transcription. The third method uses a more defined *in vitro* transcription system in which pure or partially pure transcription factors are mixed together. The first two methods result in an extract that has greatly reduced levels of P-TEFb and in the third method one of the required fractions would be P-TEFb.

10

In all cases the depleted extracts or fractions not containing P-TEFb are examined for their ability to allow Tat transactivation. A pulse-chase or continuous labeling protocol is used in which the template DNA is preincubated with the extract, 20 mM HEPES, pH 7.6 and 7 mM MgCl₂ for 20 min at 30°C. For the pulse chase protocol nucleotides, including [α -³²P]CTP, are added to start the pulse; 2 min later, excess cold CTP is added to initiate the chase. Reactions are stopped after various chase times. During the pulse, 10 μ l reaction mixtures contain the following:

20 mM HEPES (pH 7.6), 7 mM MgCl₂, 600 μ M each GTP, UTP and ATP, 5 μ Ci of [α -³²P]CTP (~1 μ M CTP), 66 mM KCl, 10-40 μ g/ml DNA template, and 3 μ l of HeLa-
FT or HeLa nuclear extract as a control.

For the chase, unlabeled CTP was added to bring the total concentration of CTP to 1.2 mM and the final reaction volume to 12 μ l. Human P-TEFb and Tat are added to the preincubation mixture where appropriate. For the continuous labeling protocol the pulse mix is 25 supplemented with 30 μ M CTP and the reactions are allowed to continue for 20 min with all combinations of inclusion of tat and P-TEFb. Reactions are stopped by adding 200 μ l of stop solution (1% Sarkosyl, 50 mM Tris, pH 8.0, 50 mM EDTA, 100 mM NaCl and 100 μ g/ml tRNA). The reaction mixtures are phenol extracted and the nucleic acids were ethanol precipitated, washed with 70% ethanol, dried and analyzed by gel electrophoresis.

To correlate the inhibition of binding to the inhibition of Tat transactivation, reactions would be supplemented with increasing concentrations of candidate small compounds or proteins. The inhibition would be quantitated by measuring the amount of runoff transcript or
5 the amount of bound P-TEFb and activity plotted versus the amount of the inhibitor. 50% inhibition points would be determined and compared in the two assays. If inhibition was similar for a given compound in both assays it would be concluded that the compound functioned by inhibiting the interaction of P-TEFb with tat.

10 **B. Results**

Control studies indicated that P-TEFb binding was specific. A GST-Tat protein containing a single amino acid change in the activation domain that causes Tat to lose its ability to transactivate *in vivo* did not bind P-TEFb. Other controls, glutathione beads alone or beads with GST only, also did not bind P-TEFb. These results indicate that P-TEFb is the CTD kinase
15 that binds to Tat. The ability of tat to increase the elongation potential of RNA polymerase II can now be explained through its interaction with P-TEFb.

Further conformation that the Tat/P-TEFb interaction is important is obtained by determining that human P-TEFb is one of the factors required for reconstructing tat transactivation *in vitro*. A protein blot containing all the required factors is probed with the human P-TEFb small subunit antibodies to determine which fraction contains P-TEFb. Since it
20 is known that the human P-TEFb elutes in the high salt step from phosphocellulose, required factors from that region are likely candidates. Confirmation of the Tat/P-TEFb interaction is obtained by correlating results with inhibitors of the binding assay (microtitre plate assay) with
25 the *in vitro* transcription assays.

EXAMPLE 7
P-TEFb is Required for HIV-1 Tat Transactivation *in vitro*

A. Generation of Antibodies of PITALRE

5 PITALRE-CT antibodies were affinity-purified rabbit IgG directed to the C-terminal 20 amino acids of PITALRE (Santa Cruz Biotechnology). Antibodies against the whole PITALRE were generated using purified recombinant protein as antigen. First two primers
10 5'-pGCAGGATCCAGAATTCCATATGGCAAAGCAGTACGACTCGG-3' (SEQ ID NO:41) and 5'- pCAGTACTCGAGTTATCAGAAGACGCGCTCAAAC - 3' (SEQ ID NO:42) were used in a PCR™ reaction to amplify the cDNA of human P-TEFb small subunit. The human brain cDNA mix (Clontech Co.) was used as template. The PCR™ product was digested with *Eco RI* and *Xho I*. The resulting 1.1 kb fragment was purified using Qiagen Gel Extraction Kit™ (Qiagen Co.). pET21a (Novagen Co.) was digested with *Eco RI* and *Xho I*. The resulting 5.4 kb fragment was purified using Qiagen Gel Extraction kit™ (Qiagen Co.). The 1.1 kb fragment and 15 5.4 kb fragment were ligated by using T4 DNA ligase. After amplification, the cloned vector was digested with *Nde I* and the larger fragment (6.5 kb) was purified using Qiagen Gel Extraction kit™ (Qiagen Co.) and religated. The final vector was amplified and transformed into DE3 (BL21) competent cells for expression of human P-TEFb small subunit.

20 The transformed DE3 cells were grown to OD₆₀₀ = 0.6 and induced with 1 mM IPTG. After a 3 hour induction, the cells were collected and lysed by passing through a French press three times. The lysate was subjected to centrifugation at 15,000 × g for 30 minutes. The pellet was solubilized in 0.1 M TUS (20 mM Tris, pH7.5, 0.1 M NaCl and 7M Urea) and loaded onto a Mono Q column (Pharmacia Co.). The flow through (FT) fraction of the Mono Q column was 25 loaded onto a Mono S column (Pharmacia Co.). The flow-through fraction of the Mono S column was subjected to dialysis against Phosphate buffer (20mM phosphate, pH7.0). The dialyzed solution was centrifuged at 15,000 × g for 30 minutes. The pellet was suspended in Phosphate Buffer (20mM phosphate, pH7.0) and used to generate rabbit antibodies following standard protocols (Pocono Rabbit Farm and Laboratory, Inc.). Preimmune serum was obtained 30 before injection of antigen (the human P-TEFb small subunit) into the rabbit. Test bleeding was

performed 42 days after the first injection of the antigen and antisera were generated monthly after the test bleeding.

B. Immunodepletion of human P-TEFb

5 Immunodepletion was performed by passing Protein A Sepharose-precleared HNE (in 20 mM HEPES, pH 7.6, 15% glycerol, 165 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.1 mM PMSF) through two affinity columns made with Protein A Sepharose beads pre-bound with anti-PITALRE-CT antibodies or control IgGs (affinity-purified rabbit anti-goat IgG, Sigma). For every 50 µl of HNE, 10 µl of protein A beads containing 1 µg of bound IgG was used. After
10 depletion of HNE, the antibody-containing beads were extensively washed with 100 times the bead volume of 20 mM HEPES (pH 7.6), 0.5% NP-40, 1% Triton X-100, and 5 mM DTT, 600 mM or indicated concentration of NaCl, and then washed with 25 times the bead volume of 20 mM HEPES (pH 7.6) and 1 mM DTT. The amounts of the washed beads used for the kinase assays and silver staining were the equivalent of 1 µl and 10 µl of HNE, respectively.

C. CTD kinase assay

15 CTD kinase assays were performed in a 20 µl reaction containing 20 mM HEPES (pH7.6), 10 µM ATP, 5 µCi [γ -³²P]ATP, ~10 ng *Drosophila* RNA polymerase II, 5 mM MgCl₂, and purified *Drosophila* P-TEFb or immunoprecipitated human P-TEFb. In the reactions containing DRB it was added to 50 µM unless other indicated amounts were used. The reactions were incubated at 30°C for 1 hour.

D. TAK activity assay

25 Preparation of Tat fusion proteins and the TAK pull-down assay were conducted as described by Herrmann and Rice (Herrmann and Rice, 1993; 1995) with modifications as described. Nuclear extract, cytoplasmic extract, or fractions containing partially purified TAK were incubated with glutathione-Sepharose beads containing GST-Tat fusion proteins for 1 hour at 4°C with gentle rocking. DE3 (BL21) bacteria containing the GST-Tat expression vectors were obtained from NIH AIDS Research and Reference Reagent Program. For maximum

sensitivity, GST-Tat48 Δ was used unless otherwise specified. The beads (30 μ l, 50% slurry) were washed 6 to 8 times with 1 ml EBCD buffer (50 mM Tris, pH 8.0, 120 mM NaCl, 0.5% Nonidet P-40, and 5 mM DTT) containing 0.03% SDS, then 2 to 4 times with Tat kinase buffer (TKB/Mg: 50 mM Tris-HCl, pH 7.6, 5 mM DTT, 5 mM MnCl₂, and 4 mM MgCl₂), and brought 5 to 50 μ l with TKB/Mg buffer and kinase assay mix. The final reaction contained 2 μ M ATP, 10 μ Ci γ -³²P-ATP (ICN, 3000 Ci/mmmole), and 50-100 μ M of CTD trimer peptide CTD3 (ACSYSPTSPSYSPSPTSPSYSPSKK, SEQ ID NO:68). Reactions were incubated at 25°C for 40 minutes, stopped by boiling in Laemmli sample buffer, and resolved by electrophoresis in 15% polyacrylamide: bis-acrylamide (30%:0.15%) gels.

10

E. Partial Purification of TAK

TAK was partially purified from a HeLa cell cytoplasmic fraction (2.6 gm protein) prepared according to Ausubel *et al.* (Ausubel *et al.*, 1989) except that dialysis was omitted. Proteins precipitating between 10% and 40% saturation of ammonium sulfate (816 mg) were resuspended in DEAE buffer (25 mM HEPES, pH 7.6, 150 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.1 mM PMSF, 4 mM MgCl₂, protease inhibitors aprotinin, leupeptin, pepstatin A at 1 μ g/ml each, and 10% glycerol), dialyzed against the same buffer, and applied to a 230 ml DEAE-Sepharose column equilibrated in DEAE buffer. The flow-through fraction (350 mg protein) was concentrated by 50% ammonium sulfate precipitation. The proteins were resuspended in HE buffer (same as DEAE buffer except 25 mM HEPES, pH 6.9, and 100 mM KCl were used) and loaded onto a perfusion chromatography heparin affinity column (POROS 20 HE; PerSeptive Biosystems). The column was washed with the same buffer and proteins were eluted with a linear gradient of 100-500 mM KCl in 15 column volumes of HE buffer. TAK activity eluted from 200 mM to 250 mM KCl. Active fractions were diluted with an equal volume of the same 15 buffer except that the pH was 8.0 and KCl was omitted, and loaded onto a perfusion chromatography cation exchange column (POROS 20 SP; PerSeptive Biosystems). The column was washed with SP buffer (the same as HE buffer except that the buffer was 25 mM HEPES pH 7.5) and eluted with a linear gradient of 100-500 mM KCl in 15 column volumes of SP buffer. The active fractions (1 mg protein) eluted at ~300 mM KCl.

F. Transcription assay

The transcription template (*Nco* I-digested pLTR-4/CAT vector containing the HIV-1 LTR from -153 to +80) was used in a pulse-chase transcription experiment. Reactions (12 µl total; 20 mM HEPES, pH 7.6, 7 mM MgCl₂, 20 µg/ml DNA template, 3 µl HNE, 64 mM KCl, with 50 µM DRB or 0.3 µl *Drosophila* P-TEFb (Marshall *et al.*, 1996) as indicated) were pre-incubated for 20 minutes at 30°C, pulsed for 2 minutes by adding ATP, GTP, UTP to 600 µM each and 5 µCi [α -³²P]CTP (~0.1 mM), and chased for 5 minutes by adding CTP to 1.2 mM. Reactions were stopped, phenol extracted and analyzed on a 6% polyacrylamide gel as described (Marshall and Price, 1995). pLTR-TAR-Luc and pLTR-DTAR-Luc plasmids were digested with *Eco* RI to generate TAR and DTAR template that were used in the studies. The TAR template contains HIV-1 LTR from -475 to +76, while the DTAR template contains HIV-1 LTR from -475 to +19. Transcription with TAR and DTAR templates generates runoffs of 694 and 646 nucleotides, respectively. The transcription reaction mix contained 20 mM HEPES (pH7.6), 7 mM MgCl₂, 60 mM KCl, 0.5 µl HNE (or depHNE as indicated), 20 µg/ml TAR template (or DTAR template as indicated), indicated amount of HIV-1 Tat (86 residue HIV-1 Tat followed by a streptavidin binding tag at its carboxyl terminus) and DRB. The reaction mix was pre-incubated for 15 minutes at 30°C and the transcription was started by adding nucleotides to final concentrations of 50 µM ATP, 50 µM GTP, 50 µM UTP, 10 µM CTP, and 5 µCi [α -³²P]CTP (~0.1 mM), and continuously labeled for 20 minutes, or by adding nucleotides to final concentrations of 50 µM ATP, 50 µM GTP, 50 µM UTP, and 5 µCi [α -³²P]CTP (~0.1 mM), and pulsed for 2 minutes. The reactions were stopped and reaction mix were phenol-extracted. The transcripts were then analyzed on a 6% polyacrylamide gel as described (Marshall and Price, 1995).

25

G. PITALRE is a Component of Human P-TEFb

According to the present discovery, if PITALRE is the functional homologue of the small subunit of *Drosophila* P-TEFb. Therefore, removal of PITALRE from HeLa nuclear extract (HNE) should eliminate DRB-sensitive runoff transcripts. To confirm this, PITALRE was

immunodepleted from HNE with antibodies directed against the last 20 amino acids of PITALRE which are not shared with other known kinases. Western blot analysis indicated that PITALRE was removed by anti-PITALRE antibodies to levels below detection, but not by control antibodies. The depleted HNE was unable to generate DRB-sensitive 633-nt run-off transcripts
5 from an HIV-1 LTR template in a pulse/chase transcription reaction. Addition of pure *Drosophila* P-TEFb to the depleted extract restored DRB-sensitive transcription. These results indicate that depletion of PITALRE abolished human P-TEFb activity, supporting the hypothesis that PITALRE is a component of human P-TEFb.

PITALRE was not previously known to be a CTD kinase. Therefore, the material
10 immunoprecipitated was examined during the depletion of HNE for CTD kinase activity. The antibody-loaded beads containing PITALRE were washed extensively with high salt and subjected to a CTD kinase assay. Similar to *Drosophila* P-TEFb, beads containing PITALRE (together with any other strongly associated proteins) were able to convert the largest subunit of *Drosophila* RNA polymerase II to the hyperphosphorylated I_{Io} form. Control beads were
15 inactive. As expected, all phosphorylation was sensitive to 50 mM DRB. In the control reaction with *Drosophila* P-TEFb autophosphorylation of both subunits (43 and 124 kDa) was seen. In the reaction with beads containing PITALRE antibodies, the 40-kDa PITALRE, a band of similar size to the large subunit of *Drosophila* P-TEFb and several other bands were phosphorylated.

To examine the association of other proteins with PITALRE, immunoprecipitates were
20 washed with buffer containing non-ionic detergents and increasing amounts of salt. The proteins associated with the beads were analyzed by SDS polyacrylamide gel electrophoresis followed by silver staining. When no salt was present many proteins were retained on the beads. Most of these proteins were removed by washing with 200 mM NaCl. No changes occurred in the proteins visible after a wash with buffer containing NaCl higher than 400 mM. Besides
25 immunoglobulin heavy and light chain and PITALRE (40 kDa), proteins with sizes 87, 105, 133, and 140 kDa were found. When rabbit anti-goat IgG control beads were used no other proteins except for the immunoglobulins were seen after high salt washes. The immunoprecipitates were incubated with [γ -³²P]-ATP to determine which proteins became phosphorylated. The beads

washed without salt carried out extensive phosphorylation of many protein substrates with little DRB-sensitivity. After washing with 400 mM or higher concentration of NaCl, only a few proteins were labeled and all phosphorylation was DRB-sensitive. Of the major proteins associated with PITALRE only the 105 kDa protein was not phosphorylated. At all salt concentrations the partially DRB-sensitive phosphorylation of a 207 kDa protein was observed. Except for PITALRE the identity of the other proteins is unknown. The sizes of these proteins do not correlate with subunits of other known basal transcription factors. The other proteins could be constituents of a larger complex containing P-TEFb or different complexes containing PITALRE.

A distinguishing characteristic of P-TEFb is its sensitivity to the kinase inhibitors DRB and H-8. *In vitro* transcription and the CTD kinase activity of *Drosophila* P-TEFb are both inhibited by these two compounds and in both assays DRB is 10 fold more potent than H-8 (Marshall and Price, 1995; Marshall et al., 1996). The effect of DRB and H-8 were determined on transcription in the HNE (FIG. 7A) and on kinase assays using the immunoprecipitated human P-TEFb (FIG. 7B). In both assays DRB was the more effective inhibitor. DRB and H8 compete with ATP for binding to the kinase active site and, therefore, it is inappropriate to compare directly the 50% inhibition points under different conditions, especially if different concentrations of ATP are used (Marshall et al., 1996). However the ratio of 50% points for different compounds under identical conditions can be compared. This ratio is more likely to be independent of assay condition. The ratio of 50% inhibition points (H-8/DRB) was $23 \mu\text{M} / 1 \mu\text{M} = 23$ for the transcription assay and $16 \mu\text{M} / 0.65 \mu\text{M} = 25$ for the CTD kinase assay. These ratios were similar to each other suggesting that the same kinase, namely human P-TEFb, was inhibited in both assays. Considering the sequence similarity between the small subunit of *Drosophila* P-TEFb and PITALRE, the functional similarity between *Drosophila* P-TEFb and the activity removed from HNE by PITALRE antibodies, and the presence of other potential subunits in the immunoprecipitates, the inventor concluded that PITALRE is a component of human P-TEFb.

H. P-TEFb Specifically Associates with the Activation Domain of HIV Tat

Several lines of evidence led the inventor to believe that P-TEFb associates with the viral transactivator Tat. Tat transactivation is sensitive to DRB (Braddock et al., 1991; Marciniak and Sharp, 1991) and requires the CTD (Chun and Jeang, 1996; Parada and Roeder, 1996; Yang et al., 1996), and Tat associates with a DRB sensitive CTD kinase (TAK) (Herrmann and Rice, 1993; Herrmann and Rice, 1995; Chun and Jeang, 1996; Yang et al., 1996). The hypothesis was tested by ascertaining if the human P-TEFb kinase associates with Tat during incubation with HeLa extracts. Glutathione beads containing various GST-Tat fusion proteins were incubated with HeLa extract and extensively washed. Proteins associated with Tat constructs containing an intact activation domain (Tat72 and Tat48D) were able to phosphorylate the synthetic peptide, CTD3, as well as RNA polymerase II. GST-Tat fusions containing mutations in the activation domain that abolish Tat-transactivation (Herrmann and Rice, 1995; Rice and Carlotti, 1990) were not able to pull down TAK. The proteins associated with the Tat constructs were probed with anti-PITALRE antibody by western blot analysis. PITALRE was detected only when the constructs contained an intact Tat-transactivation domain. This indicates that P-TEFb is a Tat associated CTD kinase.

When human P-TEFb was depleted from HNE by antibodies to PITALRE, TAK activity (assayed by GST-Tat48 Δ pull-down) was reduced to less than 2% of that found in the intact extract. This strongly suggests that under the conditions used human P-TEFb is the predominant CTD kinase that associates with Tat. Others have shown that TFIIH can associate with Tat (Parada and Roeder, 1996; Garcia-Martinez et al., 1997b), but under the conditions used no p62 subunit of TFIIH was detected as being bound to Tat nor did the inventor detect a reduction in the amount of the subunit in the PITALRE-depleted extract. In addition, when glutathione beads containing TAK were probed with antibodies to all three subunits of CAK (CDK7, cyclin H, MAT1; supplied by D Morgan) no evidence of the TFIIH associated kinase was found.

To further confirm that human P-TEFb can associate with Tat, TAK was partially purified from HeLa cells by sequential chromatography on DEAE, Heparin and SP resins. Fractions eluted from the second and third columns, Heparin and SP, were assayed for TAK

activity and probed for human P-TEFb. In the eluate of both columns, TAK activity assayed by GST-Tat48D pull-down correlated with P-TEFb determined by western analysis using two different preparations of PITALRE antibodies. A GST construct with a mutation in the Tat activation domain did not become associated with CTD kinase activity when incubated with the same column fractions. In addition, the drug-sensitivity of TAK and human P-TEFb were compared. As expected, TAK activity was inhibited by DRB and H-8 in a manner similar to P-TEFb.

I. Human P-TEFb is Required for Tat-Stimulated Elongation

To investigate the function of P-TEFb in Tat transactivation directly, the effect of Tat on initiation and elongation in whole or PITALRE-depleted extracts were compared (FIG. 8). Two different DNA templates were used. The TAR template contained HIV-1 LTR sequences from -475 to +76 while the DTAR template lacked the sequence encoding TAR and contained HIV-1 LTR sequences from -475 to +19. Using HNE (not depleted) and a continuous labeling protocol, 10 ng/ μ l Tat stimulated the generation of runoff about 12 fold (filled bars in FIG. 8). This effect of Tat was mostly inhibited by 50 μ M DRB and required the TAR sequence. A DRB titration indicated that the DRB sensitivity of Tat-stimulated runoff was similar to that determined for the low level of runoff from the HIV-LTR in the absence of Tat. When PITALRE-depleted extract was used, the majority of the stimulatory effect of Tat was abolished. This result strongly suggests that human P-TEFb is required for efficient transactivation by Tat. Add-back of *Drosophila* P-TEFb to the depleted extract stimulated runoff in the presence or absence of Tat, but did not restore the ability of Tat to specifically enhance elongation. This suggests that other required factors were removed by the depletion of PITALRE or that *Drosophila* P-TEFb lacks appropriate domains required for interaction with Tat or other cofactors. Addition of the high salt washed PITALRE immunoprecipitate had no effect on runoff in the presence or absence of Tat, indicating that immobilization had a negative impact on the function of P-TEFb. Although consistent with an effect on elongation, the experiment did not rule out the possibility that the major effect of Tat was on initiation.

From the autoradiograph and the normalized quantitation data graphed in FIG. 8 (filled bar) it was evident that Tat had a modest effect on transcription (about two fold increase of runoff) in the presence of 50 μ M DRB, in the presence of DTAR template, or even in the absence of human P-TEFb. To further analyze the effect of Tat on transcription, a similar sets of studies 5 was performed, except that a pulse-only transcription protocol was used instead of a continuous labeling protocol. Quantitation of the short pulsed transcripts (less than 20 nucleotides in length) gives the efficiency of initiation. Under these conditions it was clear that increasing Tat had the effect of increasing transcription initiation (open bars in FIG. 8). At the highest level of Tat used 10 (10 ng/ μ l) initiation increased about two fold compared to when no Tat was added. This effect on initiation did not require TAR (FIG. 8). Moreover, the effect of Tat on initiation was not sensitive to DRB and did not require P-TEFb. The effect of Tat on initiation can be used to explain the slight effect of Tat on the generation of runoff in the presence of DRB, DTAR template, or depleted HNE. The two assays taken together indicate that the major effect of Tat is 15 on elongation, although initiation is slightly affected. Most importantly, P-TEFb is required for the effect of Tat on elongation, but not its effect on initiation.

EXAMPLE 8 **Determining the Region of P-TEFb Responsible for the Interaction with Tat**

Although Example 7 clearly shows that P-TEFb associates with the HIV protein Tat, the interaction of Tat with P-TEFb could be through either subunit. A combination of *in vivo* and *in vitro* assays are used to localize the interaction surface. Both subunits are expressed, *e.g.* in *E. coli*, purified and the recombinant proteins used *in vitro* to determine which one or both of the subunits bind to Tat or compete for the binding of intact P-TEFb to Tat. The peptide domains 20 are also expressible as GST fusion proteins, and soluble proteins are obtained by removing the GST moiety by specific protease digestion.

In order to identify proteins that bind to P-TEFb, the inventor utilizes the yeast two-hybrid system to identify proteins that associate with P-TEFb *in vivo* (Fields and Song, 1989; 30 Chien *et al.*, 1991; Durfee *et al.*, 1993; Harper *et al.*, 1993, U.S. Patent No. 5,667,973,

incorporated herein by reference). This system utilizes the yeast GAL4 protein, a well studied eukaryotic transcriptional activator protein. Transcriptional activator proteins are proteins that bind to cognate promoter elements upstream of particular genes, and thereupon activate transcription of the associated gene (Johnston, 1987). This activation function is believed to
5 occur through recruitment of specific proteins which are required, along with RNA polymerase, to effect transcription.

Reports have shown that the DNA binding function and the transcriptional activation function reside in two distinct regions of the GAL4 protein. Further, it has been shown that these
10 regions are able to be localized to relatively short peptide regions, which can function separately (Brent and Ptashne, 1986; Keegan *et al.*, 1986). This ability forms the basis of the two-hybrid system. In the yeast two-hybrid system, the gene encoding the known protein is cloned as a fusion protein with the GAL4 DNA binding domain. Then a gene encoding a protein suspected of interacting with the known protein, or a cDNA library (to assay for the presence of a gene
15 encoding an interacting protein) is cloned as a fusion protein with the transcriptional activation domain. Both constructs are then introduced into individual yeast cells.

Binding *in vivo* between the two fusion proteins recombines the GAL4 DNA binding domain with the GAL4 transcriptional activation domain, which leads to the transcriptional activation of a marker gene operatively positioned downstream of a GAL4 binding site. The transcriptional activation construct is then recovered from the identified yeast cell, and the interacting protein is identified by DNA sequencing.
20

The standard yeast two-hybrid system described above is a positive association screening
25 assay. A variation of the two-hybrid assay is the subject of U.S. Patent number 5,525,490, by Erickson and Powers, issued June 11, 1996 and incorporated herein by reference. Using this reverse two-hybrid system, inhibitors of protein-protein interaction can be assayed. In this system, the interaction between the fusion proteins represses the transcription of a marker gene. Agents (for example, but not limited to, proteins) which disrupt the association of the fusion
30 proteins are identified by the increase in transcription of the marker gene.

The expressed proteins are included in binding reactions containing GST-Tat immobilized on glutathione beads and P-TEFb. In the event that additional cofactors are required, they are supplied as a crude extract or partially purified fraction. If there are cofactors
5 required for the binding of P-TEFb to Tat, the binding assay with pure P-TEFb can be used as a means to purify the cofactor. The cofactor can then be used as a potential target for drugs that block the function of Tat through the loss of the ability to bind to P-TEFb.

After washing the beads, the binding of P-TEFb is monitored by western analysis by
10 using antibodies to the desired subunit of human P-TEFb. The subunit that inhibits the binding of P-TEFb to Tat is a strong candidate for being the one that contains the domain that interacts between Tat and P-TEFb.

At the same time, the individual proteins are monitored directly for their ability to bind to Tat. The binding of portions of the subunits to Tat are monitored by using antibodies (see Example 5) that recognize the peptides. Antibodies from the crude antisera raised against individual subunits can be used. (Antibodies to the large subunit of P-TEFb can be used similarly.) Alternatively, if the activity of antibodies in the crude antisera is insufficient then antibodies are raised against the portions of the subunits that inhibit P-TEFb binding to Tat.
15 Once direct binding of a peptide containing the interactive domain to Tat is detected, then the identified peptide is useful for screening for drugs that inhibit the binding of P-TEFb to Tat. It is envisioned that these peptides are easy to prepare in large quantities using recombinant DNA technology for use in screening assays for inhibitors of activators of P-TEFb.
20

A complimentary approach is to utilize a reporter gene assay such as one that utilizes chloramphenicol acetyltransferase (CAT). The use of reporter gene assays are well known to those of skill in the art. Recombinant HeLa cells are prepared such that they express both Tat and the kinase and CTD domains of the subunits of human P-TEFb and are also able to support Tat-transactivation of an HIV promoter driven CAT construct (HIV-CAT). It is envisioned that
25 CAT activity is monitored when HIV-CAT is transiently transfected alone or co-transfected with

increasing amounts of different constructs that respectively express the desired domain(s) of the human subunits. A decrease in CAT activity signifies that expression of the specific domain(s) inhibits the function of Tat.

5 Antibodies to the respective subunit domains of P-TEFb that interact with Tat are generated, using standard methods as modified and described in Example 5, so that expression of the domains can be verified using western blotting. Polyclonal antibodies against the entire large subunit can also be used, but it is envisioned that the individual domains are useful as antigens for the production of more specific antibodies.

10 These *in vivo* studies are useful for the identification of regions of P-TEFb that interact with Tat or that directly affect the normal cellular function of P-TEFb. Comparison of the *in vivo* expression results with the *in vitro* binding studies generally allows correlation of binding to function and specifically allows the determination of interaction domains that specifically affect Tat-transactivation but not the general function of P-TEFb.

15 Although it is clear that P-TEFb is essential for elongation to occur, it is possible that the interaction between P-TEFb and Tat is not direct. One or more other proteins may form a bridge between P-TEFb and Tat. The identification of any such proteins would be advantageous for the understanding of Tat-transactivation and the development of drugs that inhibit its interaction. Candidates for Tat/P-TEFb bridging proteins are obtained using the yeast two hybrid system, described above, to select for proteins that bind to P-TEFb.

20 For example, the bait vector which contains a region or regions of human P-TEFb fused to a DNA binding protein is constructed and transfected into yeast cells along with a human cDNA library fused to transcription activation domain. Yeast are selected based on their ability to activate transcription of the gene downstream of the DNA binding region. The large subunit of P-TEFb is detected if the small subunit of P-TEFb is used as bait. Once constructed, this method is then useful for cloning the large subunit. After the large subunit is cloned, the whole 25 subunit or portions of it are useful as bait to find proteins that interact with P-TEFb. Candidates

for bridging proteins are screened for their ability to function *in vitro* by allowing or enhancing the binding of P-TEFb to Tat.

EXAMPLE 9
Interaction of the *Herpes* Transactivator VP16 with P-TEFb

The *Herpes* virus transactivator VP16 increases the ability of RNA polymerase II to synthesize DRB-sensitive runoff transcripts *in vitro*. Most promoters have sites for activator proteins that are present in cell extracts and can allow for the production of DRB-sensitive long transcripts. But under transcription conditions like those in Marshall and Price (1992), the basal adenovirus E4 promoter contains no activator binding sites except for 5 Gal4 binding sites which give rise to very low levels of DRB-sensitive runoff transcripts in a *Drosophila* K_c cell nuclear extract. Reasonable levels of initiation occur but most of the transcripts (>95%) are short (abortive).

When a protein, containing the N-terminal portion of the VP16 activation domain (amino acids 413-456) fused to the DNA binding region of Gal4 (amino acids 1-147), was added to the transcription reaction the amount of DRB-sensitive runoff increased 6.2 fold. Only part of the increased production could have been due to increased initiation because only 1.8 times as many polymerases initiated in the presence of Gal4-VP16 (as determined by quantitating the transcripts generated during a short pulse).

Proteins used to fuse to the DNA binding region were Gal4-VP16 N (wildtype) and Gal4-VP16 N-FA442 (mutant) which has the single amino acid change indicated. The proteins were made by inserting the indicated sequences downstream of the P_{tac} promoter in an expression vector and transforming *E. coli* XA90 cells. 1 liter of cells were grown in LB +Amp to 0.7 OD₆₀₀. After a 3 h. induction with IPTG cells were lysed by sonication in 80 ml of lysis buffer (20 mM HEPES, pH 7.5, 10 µM zinc acetate, 20 mM β-mercaptoethanol and 200 mM NaCl). The proteins were precipitated with 0.25% Polyethyleneimine, spun, and suspended in lysis buffer. The pellet was redissolved in lysis buffer with 750 mM NaCl and spun again. The

supernatant was ammonium sulfate precipitated with 40% saturated solution. The pellet was redissolved in 20 ml lysis buffer with 100 mM NaCl and 1 mM DTT instead of β -mercaptoethanol. The supernatant was dialyzed against lysis buffer with 100 mM NaCl.

5 Transcription conditions were as used in Marshall and Price (1995) and as described in Example 1. If indicated 75 ng of Gal4-VP16 N or Gal4-VP16 N-FA442 were added per reaction.

10 A mutant Gal4-VP16 that does not activate well *in vivo* stimulated DRB-sensitive runoff transcripts 2.8 fold and had a slight stimulatory effect on initiation (1.5 fold). Since P-TEFb is responsible for DRB-sensitive runoff transcripts these data strongly indicate that P-TEFb binds to the VP16 activation domain.

15 A Packard InstantImagerTM was used for the quantitation. After running the products of the transcription reaction on a denaturing gel, the gel was dried and imaged for 30 min in the InstantImagerTM. Initiation was compared with and without the activator protein by comparing the counts generated by all transcripts during the pulse or by comparing the counts that reached runoff after chasing with 250 mM KCl which has been shown to be directly related to initiation. Both methods gave the same results. Radioactivity in the region of runoff after a normal low salt chase was used to quantitate productive elongation products. DRB-sensitive runoff was calculated by subtracting the runoff in the lane with DRB from the runoff in the lane without DRB. P-TEFb is present in the Kc cell extract used in the reactions. 75 ng of Gal4-VP16 N or N mutant was used in each 12 μ l reaction.

20 These results were extended by removing the upstream region of the *Drosophila* actin template normally used in K_c cell extract transcriptions and inserting the Gal4 binding sites (pBAG1). Transcription of 20 μ g/ml pBAG1 using a normal pulse chase conditions gave rise to a low level of DRB-sensitive runoff transcript but inclusion of 75 ng of Gal4-VP16 N increased initiation slightly and significantly stimulated elongation. These are the same results obtained when using the E4 promoter. Therefore, the effect of VP16 is independent of the basal promoter

that is used. In addition elongation was stimulated by Gal4-VP16 N from the E4 promoter in HeLa nuclear extract transcriptions.

EXAMPLE 10
Assay for Drugs that Block the Interaction of P-TEFb and Tat

A major goal in medicine today is to find drugs that can inhibit and prevent the occurrence of AIDS and spread of HIV infection. It is envisioned that because P-TEFb has a central role in cell proliferation and interacts with Tat, P-TEFb is useful as a tool in screening immunoassays to find therapeutic drugs that will inhibit or block the proliferation of HIV.

Immunoassays, such as an ELISA, are useful to screen for drugs that inhibit the interaction of Tat with P-TEFb. For example, Tat-coated microtiter plates are incubated with pure P-TEFb or P-TEFb and any other required factors (such as human nuclear extract or fractions derived there from) needed to cause association with Tat. As indicated previously, a requirement for additional cofactors is unlikely, but the existence of needed cofactors does not complicate this assay when HeLa extract is used. As before these cofactors can be supplied by using the crude cellular extract.

As the interaction between components in this assay are interdependent upon the concentration and activity of each other, parameters for each component are established by initially using wide ranges and optimizing the amount of each component as desired. For example, 1-10 µg of Tat can be bound per well initially, and 15 µl of HeLa extract provides sufficient P-TEFb (about 10 ng) to establish binding. Primary anti-P-TEFb small subunit antibodies can initially be in a 1:1000 dilution of antiserum. Secondary antibodies can be initially diluted 1:20,000. Of course, these amounts, as with all other parameters, may be optimized as desired.

It is envisioned that any peptides, proteins or polypeptides derived from P-TEFb and shown to be active towards Tat may be used as a P-TEFb template in an immunoassay. After

washing, the plates are incubate with antibodies that have been prepared using standard protocols and methods as exemplified in Example 5 and that recognize either the small or large subunit, intact P-TEFb or peptides derived there from (see Example 8). Following standard methods for immunoassay detection, a label is applied to the plates such that microtiter wells that contain
5 P-TEFb bound to Tat yield a positive signal.

To determine that a candidate drug inhibits the interaction of the P-TEFb with Tat, a competitive assay is performed. It is contemplated that P-TEFb proteins, subunits, polypeptides or peptides can be used. Decreasing or increasing concentrations of candidate drugs are added to
10 microtiter plates which contain constant amount of P-TEFb template and Tat. Disruption of the interaction of P-TEFb with Tat is detected as the reduction of the positive signal which indicates P-TEFb template bound to Tat.

Drugs which can strongly inhibit or even prevent the interaction of Tat and P-TEFb template are potential candidates for further screenings to determine if they interfere with normal cellular processes. Preferably, inhibitors are small compounds suitable for use as anti-HIV drugs or compounds. The small compounds would bind (like allosteric effectors do) to regions of the Tat or P-TEFb that are not involved directly in enzymatic activity and this binding would inhibit the interaction between the two proteins.
15
20

Candidate drugs are examined, for example, for their ability to inhibit the normal interaction of P-TEFb with RNA polymerase II. The ideal candidate drug only inhibits or prevents the interaction of Tat and does not inhibit or prevent the normal interaction of P-TEFb with RNA polymerase II or any other normal cellular activities. Suitable screening methods
25 include, but are not limited to, additional immunoassays, cell culture base assays, animal testing models and clinical trials.

EXAMPLE 11
Characterization of Human P-TEFb.

5 Immunoprecipitation of PITALRE from HeLa extracts indicates that several other proteins are tightly associated with the kinase subunit even in the presence of 1 M NaCl and non-ionic detergents. Proteins of molecular masses, 87, 105, 133, 140 and 207 kDa are visible with silver staining, but the stoichiometry of these proteins is not known. There could also be additional proteins that do not stain well or that are hidden under the immunoglobulin bands present in the immunoprecipitates.

10

A. Characterization of the Two Potential Cyclin Subunits of Human P-TEFb.

15 Although northern analysis indicates that the intron containing mRNA is a minor species it is not known if the protein encoded becomes functionally associated with PITALRE. To examine the association of the proteins encoded by these potential cyclin subunit clones antibodies are made to each in rabbit (Pocono Rabbit Farm). These antibodies are used to probe western blots of PITALRE immunoprecipitates from HeLa cell extract. The recognition of proteins in PITALRE immunoprecipitates is strong evidence for a functional interaction. If the potential cyclin subunit antibodies recognize proteins associated with PITALRE, the size of these proteins is compared with the size of the proteins expressed in the baculovirus system. If there is a discrepancy between the size of the endogenous proteins and the recombinant proteins additional clones are obtained that encode a protein of appropriate size. Results indicate that 5' 20 and 3' RACE products frequently have unusual sequences due to the presence or absence of intronic sequences and it is possible that the clones obtained so far may not have the correct amino- or carboxyl-terminal ends. However, this evidently does not negate their function or the ability of the skilled researcher to identify correct terminal sequences. The protein encoded by human HBL3 cDNA was co-expressed with PITALRE in a baculovirus system and produced an active heterodimeric kinase. The protein encoded by human cyclin HBL3 exhibited a mobility identical to the 87 kDa protein found in PITALRE immunoprecipitates. These results suggest 25 that the protein subunits described herein are P-TEFb. Further confirmation is obtained by using antibodies to the cyclin subunit to probe PITALRE immunoprecipitates..

To determine if the identified cyclin protein is part of an active complex immunodepletions like those done with PITALRE (Example 6) are performed. If the ability to synthesize DRB-sensitive transcripts is eliminated from HeLa extracts by the antibodies a strong
5 functional connection is made between the cyclin subunit and P-TEFb. At the same time antibodies against the identified cyclin subunit are used to carry out immunoprecipitation and the resulting pattern of proteins are compared to PITALRE immunoprecipitates. Besides PITALRE, it is important to determine what other proteins the two immunoprecipitates have in common.
10 The current understanding of complexes containing PITALRE does not make clear if all the observed proteins are in a single complex or if perhaps there are several different PITALRE containing complexes. Using antibodies against another protein in the complex allows a better understanding of what proteins are in the complex with functionally active PITALRE.

B. Analysis of Complexes Containing PITALRE and a Cyclin Subunit of P-TEFb.

Since P-TEFb is a cyclin dependent kinase it is possible that it may associate with other proteins that regulate its activity. Many CDKs can be inhibited by small molecular mass proteins that bind to the cyclin/kinase pair (Morgan, 1995a; MacLachlan *et al.*, 1995). If such an inhibitory protein binds to P-TEFb, the complex would be inactive and would be difficult to find with conventional assays that require activity. Armed with antibodies against several subunits of human P-TEFb different complexes containing PITALRE are examined to determine whether they have activity or not. HeLa nuclear extract are fractionated using conventional chromatographic methods and complexes containing PITALRE or the cyclin subunit are located by western blot. Initially, proteins in the extract are bound to phosphocellulose to remove most nucleic acid and the material eluted at 1 M KCl is applied to a Sepharose S200 column to effect a
15 size separation. PITALRE and or cyclin containing complexes are identified and subjected to further purification. Immunoprecipitation is used to analyze the proteins associated with PITALRE at all stages of the purification. The column fractions and immunoprecipitates are examined for kinase activity and the column fractions are tested for their ability to function during Tat-transactivation when added back to PITALRE depleted extract. Both active and
20 inactive complexes are identified by this method. Purification of the complexes with known
25

properties allows identification of other potentially regulatory subunits through protein sequencing methods like those used in the cloning of the subunits of *Drosophila* P-TEFb.

C. Molecular analysis of human P-TEFb.

5 The availability of antibody and cDNA probes allows the assessment of expression patterns of P-TEFb. Some characterization of PITALRE has been carried out and it has been reported that PITALRE was localized to the nucleus and PITALRE mRNA was expressed ubiquitously over a variety of human tissues (Grana *et al.*, 1994). In addition, PITALRE has been mapped to a region of chromosome 9 involved in certain types of cancer (Bullrich *et al.*,
10 1995). The inventor's work indicates that mRNAs encoding both of the cyclin subunits that have been isolated are expressed in a similar manner to that seen for PITALRE. Northern blot analyses of tissues involved in the immune response are used to determine if cells that are normally infected with HIV preferentially express a specific cyclin subunit or a specific form of one the subunits.

15 Mapping of the cytogenic location of cyclin subunit(s) found to complex with PITALRE is used to determine if the areas map to other loci involved in similar types of cancer as found for PITALRE. Mapping is accomplished by using PCR primers that give specific products for human versus hamster DNA. Any correlations that are found are then useful for determining if there are genetic abnormalities in affected individuals.
20

D. Functional Analysis of P-TEFb.

An analysis of P-TEFb function *in vivo* is carried out using a co-transfection protocol with P-TEFb and an HIV-LTR reporter construct. The goal is to determine if over-expression of
25 active P-TEFb causes an increase in Tat transactivation and to determine what regions of P-TEFb are required. Both HeLa and Jurkat cell lines that support modest or very high levels of Tat transactivation respectively are used. Expression of PITALRE has been shown to modestly stimulate Tat transactivation in several cell lines even though the others have found that expression of PITALRE does not increase dramatically the level of kinase activity found in
30 PITALRE immunoprecipitates (Garriga *et al.*, 1996a). This inability to increase the kinase

activity of PITALRE is likely due to a limitation in the amount of the cyclin subunit(s) required for activity. In addition, the inventor and others (Garriga *et al.*, 1996a) have not been able to get a large over-expression of PITALRE which suggests that the level of the kinase subunit may be regulated by a post-transcription mechanism. Co-expression of a cyclin subunit may allow for greater level of expressed proteins if the ratio of the kinase/cyclin pair is involved in regulating the level of the kinase. This can be studied by examining PITALRE levels during transient expression of a cyclin subunit in an induced PITALRE cell line. If higher level of the knockout kinase are observed, it may be possible to achieve much higher levels of P-TEFb in cells that have both kinase and cyclin subunits under the control of active promoters. This type of experiment may provide general information about the regulation of PITALRE levels in the cell. If the levels of the kinase knockout PITALRE protein are increased when the cyclin subunit is expressed then the effect of the potential dominant negative mutation on cell growth may be observed. If this is the case, the induction of the PITALRE kinase knockout in the presence of over-expressed cyclin subunit may be lethal to the cells.

To examine the effect of over-expression of P-TEFb on Tat transactivation, HeLa or Jurkat cells are co-transfected with constructs that express both subunits of P-TEFb and Tat along with an HIV-LTR reporter construct. The levels of both subunits of P-TEFb are compared before and after transfection to determine the level of over-expression. Measurement of reporter activity is made with and without Tat and with and without individual subunits of P-TEFb. If over-expression of P-TEFb causes a reasonable stimulation of Tat transactivation, constructs expressing mutant P-TEFb subunits are used to assess their effects. The kinase knockout mutation of PITALRE is studied to see if it will act as a dominant negative mutation. Most importantly, the requirement for the carboxyl-terminal domain of the cyclin subunit for Tat transactivation is examined. Effects of all interesting P-TEFb constructs on a Tat responsive reporter are compared to effects on other promoters, such as the complete CMV promoter, that use different activators. These *in vivo* studies should complement similarly designed *in vitro* experiments described below.

EXAMPLE 12
Examination of the Tat/P-TEFb Interaction.

An immobilized GST fusion protein containing an intact Tat activation domain is
5 incubated with fractions around the elution point of PITALRE and extensively washed as was
done in (see Example 6). The material bound to Tat is assayed for CTD kinase activity using
RNA polymerase II as substrate (Marshall *et al.*, 1996) and for the presence of PITALRE by
immunoblotting. As was done before, negative controls are performed using Tat proteins with
mutations that neutralize the activation domain. Any complex that fractionates with the ability to
10 bind to Tat is monitored by immunoprecipitation with anti-PITALRE antibodies to identify
subunits/cofactors that are needed for association with Tat. A PITALRE containing complex that
binds to Tat may be purified in this manner and examined for its ability to support efficient Tat
transactivation. For example, the recombinant PITALRE/human cyclin HBL1-1 and
15 PITALRE/HBL3 complexes can be examined for their ability to interact with Tat and compared
to the complexes derived from fractionation of HeLa nuclear proteins.

If a purified P-TEFb complex that interacts with Tat and recombinant PITALRE/cyclin
complexes do not interact, the requirement of cofactors that do not co-purify with P-TEFb is
likely. To find such cofactors an assay in which the association of recombinant PITALRE/cyclin
20 complexes (rP-TEFb) with Tat is used to measure function and activity in the presence of
fractions of HeLa nuclear extract. The association is monitored with either a CTD kinase assay
or by immunoblot. Initially, the assay utilizes HeLa nuclear extract that was depleted of
PITALRE in the presence of 1 M KCl. This type of extract should not be depleted of factors that
interact ionically with PITALRE. An increase in the association of rP-TEFb with Tat in the
25 presence of the depleted extract suggests the presence of cofactors. If such results are obtained,
the association reactions are modified to learn if the cofactors interact with Tat or with rP-TEFb
alone. First, immobilized Tat is incubated with the depleted extract. To determine if the
cofactor(s) interact with Tat, the beads are washed and then incubated with rP-TEFb. If the
affinity of the beads for P-TEFb increases then Tat associated cofactors are suggested. A similar
30 experiment in which rP-TEFb is immobilized via its HIS-tag, incubated with the extract and then

5 incubated with Tat shows if cofactor(s) associate with P-TEFb. In the latter experiment the binding of Tat is monitored by immunoblotting with anti-Tat antibodies. It is possible that some cofactors interact with both factors, but it is also possible that a cofactor interacts with just one factor causing a conformational change that enables that factor to associate with the other partner

in the Tat/P-TEFb complex.

10 The simplest model for the function of TAR RNA is that it exerts its effect on transcription of the HIV-LTR due to its ultimate ability to recruit P-TEFb to the early elongation complex. Tat has been shown by a number of groups to bind to TAR (Fong *et al.*, 1997; Aboul-ela *et al.*, 1995; Neenhold and Rana, 1995; Rhim and Rice, 1994; Cullen, 1991) and P-TEFb or then unspecified CTD kinases have been shown to associate with Tat *in vitro* (Herrmann and Rice, 1993; Herrmann and Rice, 1995; Chun and Jeang, 1996) and *in vivo* (Yang *et al.*, 1996). However, linkage between TAR and P-TEFb or any CTD kinase has not been made.

15 TAR RNA is produced *in vitro* using T7 polymerase and is included in various binding reactions like those described above. First, it is determined if the RNA has an effect on the efficiency of interaction of P-TEFb with Tat in the crude extract. TAR RNA is titrated into binding reactions that have otherwise been optimized for association of P-TEFb to Tat. Interaction efficiency is monitored by determining the fraction of PITALRE remaining in the extract after the Tat pull-down. The assay is then performed on any promising set of fractions identified from the binding studies above. It is possible that TAR RNA may disrupt the interaction of Tat with P-TEFb. Although at first this may seem inappropriate, it is important to note that Tat may be a component of the preinitiation complex and may not need to be recruited by TAR (Garcia-Martinez *et al.*, 1997a). TAR may regulate the effect of Tat by disrupting the 20 interaction between Tat and P-TEFb, thereby allowing P-TEFb to function. What ever the results, the development of models of Tat transactivation are greatly benefit from knowing the effect of TAR RNA on the Tat/P-TEFb interaction.

25 If the results indicate that TAR enhances the association of P-TEFb to Tat then an assay in which TAR RNA is immobilized to paramagnetic particles is used. Binding of P-TEFb to the

beads dependent on the presence of Tat and crude extract or fractions generated from an extract is followed. The complexes are examined for their ability to increase the activity of associated P-TEFb. Specific activity of P-TEFb is monitored by a combination of western blotting, to determine relative amount of the factor, and CTD kinase assays, to measure the activity of the
5 factor. If the results indicate that TAR disrupts the interaction of P-TEFb and Tat, that information is useful in interpreting the outcomes of the studies described below.

EXAMPLE 13

Elucidation of the Mechanism of Tat Transactivation.

10 Progress on understanding the how Tat causes an increase in the number of polymerase molecules that synthesize full length HIV transcripts has been slow until now because of the lack of understanding of the elongation control process in general. Recent advancements in elongation control include the present identification of *Drosophila* P-TEFb as a CTD kinase (see also Marshall *et al.*, 1996), the purification and characterization of *Drosophila* factor 2, an ATP dependent RNA polymerase II termination factor (Xie and Price, 1996), and the purification of a negative human elongation factor comprised of the homologues of the yeast SPT 4 and SPT5 proteins. The present identification of the kinase subunit of human P-TEFb and the requirement for P-TEFb in Tat transactivation makes it clear that studies on the mechanism of Tat transactivation should be focused on understanding the role of P-TEFb.
15
20

A. Association of Elongation Control Factors with the Transcriptional Machinery.

The association of known elongation control factors with transcription complexes on the HIV-LTR is studied to determine the effect of Tat on this association. The present immobilized template technology (see also Marshall *et al.*, 1996; Xie and Price, 1996; Marshall and Price,
25 1992) and is used to determine if and when P-TEFb, factor 2, the SPT4/SPT5 complex, and other factors become associated with the transcription complex. Preinitiation complexes, early elongation complexes before TAR RNA is synthesized and elongation complexes containing TAR RNA are examined. The salt concentration used during washing steps and the exact
30 transcription conditions used are important factors in these studies. The strategy outlined below is given as an example of the types of studies that are done.

The overall design of the association experiments is to form preinitiation complexes on an HIV-LTR template immobilized to paramagnetic beads, allow the polymerase to progress to the appropriate stage and then wash the complexes to remove non-associated proteins.

5 Antibodies against the factors are used to probe for their presence in the isolated complexes. For this type of experiment to be successful a number of controls must be done. It is important to first determine the efficiency of initiation of preinitiation complexes because uninitiated polymerases would contribute to the signals obtained from immunoblotting of "elongation complexes". Although uninitiated polymerases could be eliminated with a 1 M salt wash, such a
10 wash would also potentially disrupt factors in a potential elongation control particle. It cannot be assumed that a pulse of NTPs allows all preinitiation complexes to initiate and form early elongation complexes. In the *Drosophila* system the inclusion of Mn⁺⁺ instead of Mg⁺⁺ in the initiation phase increased the rate of initiation and, therefore, the fraction of polymerases that initiate. Even in the presence of Mn⁺⁺ less than half of the polymerases in preinitiation complexes initiated, as evidenced by resistance to a 1 M salt wash (Marshall *et al.*, 1996). The initiation efficiency in a HeLa nuclear extract system was examined and found that Mn⁺⁺ does not stimulate the rate of initiation. This suggests that initiation is already quite efficient. To determine what fraction of polymerases initiate from the HIV-LTR in the presence and absence of Tat, immunoblotting with antibodies to RNA polymerase II is used. Preinitiation complexes are formed in the crude extract and then the complexes are washed extensively with buffer containing the highest salt possible that does not cause loss of polymerases that can initiate.
15 These washed preinitiation complexes are the starting material for an experiment to determine the efficiency of initiation. The amount of RNA polymerase II in these complexes assayed by immunoblotting are compared to the amount after initiation and after washing with 1 M salt. The
20 NTP concentrations, salt conditions and times to effect the greatest efficiency of initiation are varied as is practical. If the efficiency is greater than about 50% then the signals obtained by immunoblotting are interpretable.
25

Once good initiation conditions are established then the conditions that allow the
30 efficient formation of early elongation complexes containing nascent transcripts of between 10

and 30 nucleotides in length and complexes containing an intact TAR RNA stem and loop are determined. The shorter complexes would have made the transition from abortive initiation into elongation and, therefore, would have released contact with the promoter localized factors, but would not have synthesized TAR RNA. These conditions should be similar to those used to
5 examine initiation from the HIV-LTR . To obtain the complexes with complete TAR RNA, limiting amounts of nucleotides are used to advance the polymerase to a set of positions that allow most complexes to have RNA in the 70 to 100 nucleotide range.

Alternatively, a second method to generate complexes with RNA of more uniform length
10 is utilized such that the lactose (lac) repressor protein and a template containing the lac repressor binding site is positioned downstream of the TAR region to block elongation by the polymerase (Keen *et al.*, 1996). The length of the RNA is monitored using incorporation of α -³²P-CTP and initiation efficiency is checked as described above. If a significant number of complexes are terminating due to the action of factor 2, then antibodies are used against the newly cloned
15 human factor to deplete the factor from extracts before assembly of the preinitiation complexes. Before this is done the effect of such depletion on Tat transactivation is examined.

After finding the appropriate conditions to efficiently generate the three types of complexes, immunoblotting studies are performed with all antibodies available for the factors potentially involved in elongation control. These include, antibodies against PITALRE and the cyclin subunit(s) of P-TEFb, human factor 2, SUPT4H (Chiang *et al.*, 1996b) and SUPT5H (Chiang *et al.*, 1996a), subunits of TFIIF including those in the CAK (Morgan, 1995b), and both subunits of TFIIF (Finkelstein *et al.*, 1992). Tat is added at different times and its association with the complexes is monitored with antibodies (Brake *et al.*, 1990). Initially, elongation
20 complexes are generated in the presence of the extract and, therefore, in the presence of all factors. As the studies progress this is modified by washing the preinitiation complexes or early elongation complexes without TAR and then allowing the polymerases to advance to the next
25 complex. This allows determination of what factors are lost from earlier complexes.

Although the simplest model for Tat action involves its binding to TAR and recruitment of P-TEFb, this model is complicated by several studies that indicate that Tat can associate with RNA polymerase II and can become part of the preinitiation complex (Garcia-Martinez *et al.*, 1997a; Mavankal *et al.*, 1996; Cujec *et al.*, 1997). Using carefully controlled conditions, attempts are made to confirm that Tat associates with preinitiation complexes and extend the findings to other promoters such as the CMV promoter that is used as a control for all the experiments described above. Association of Tat with the preinitiation complex helps to explain why Tat has an effect on initiation, but does not explain the effect of Tat on elongation. It is possible that TAR RNA from an early elongation complex reaches back into the next preinitiation complex and affects the Tat RNA polymerase II interaction. It is also possible that Tat bound to TAR recruits an RNA polymerase II complex (holoenzyme) to the promoter. In either case the polymerase initiating under the influence of Tat might be more susceptible to the action of P-TEFb.

15 **B. Functional Enhancement of P-TEFb Action by Tat.**

It is clear from *in vivo* and *in vitro* studies that Tat has the ability to enhance the function of P-TEFb at the HIV-LTR. What is not clear is how this enhancement is accomplished. P-TEFb is not used only at the HIV-LTR, but rather is a cellular factor involved in elongation control at most promoters. Also the LTR does not absolutely require Tat to produce full length transcripts because if this were not the case mRNAs encoding Tat would not be produced.

One model consistent with most data is that the HIV-LTR is different compared to other promoters in that elongation complexes produced from the LTR are especially refractory to P-TEFb action. Tat may be required to eliminate a strong negative elongation potential imparted to polymerases that initiate from the HIV-LTR. This idea is supported by the effect of a number of P-TEFb inhibitors that blocked Tat transactivation. Although all of the inhibitors blocked Tat transactivation at lower concentrations than that needed to block activation by another activator on another promoter, at one log higher concentration all the inhibitors had cytotoxic effects, presumably due to inhibition of the normal function of P-TEFb. In transient transfection assays, increasing PITALRE concentration and presumably P-TEFb activity in a variety of cells was

able to stimulate reporter constructs driven by a number of promoters with different activators. While Tat transactivation on the HIV-LTR was dramatically stimulated, increasing PITALRE did not increase expression from the LTR in the absence of Tat. This supports the idea that the LTR is refractory to P-TEFb action and suggests that the increased sensitivity of Tat
5 transactivation to the inhibitors might arise due to an increase in activity that counters the action of P-TEFb.

There are several possibilities for activities that might counter P-TEFb at the HIV-LTR. One such activity might be the CTD phosphatase described by Dahmus (Dahmus, 1996; Chambers *et al.*, 1995) and others (Chambers and Kane, 1996). If this is the case an amount of inhibitor that would only partially block P-TEFb might have a large effect on the function of the factor because of the balance of the two opposing activities. Another possibility is the FBI-1 protein (Pessler *et al.*, 1997) that may be responsible for the function of the IST (inducer of short transcripts) element (Sheldon *et al.*, 1993) found in the DNA around the start point of transcription of the HIV-LTR.
10
15

The negative activity of the HIV-LTR is first produced to compare to the CMV promoter with a co-transfection/reporter system in HeLa cells and then to determine the cause of the effect. Ultimately, it is determined if the IST element is involved by examining the sensitivity of the expression of a CAT reporter activity from promoter constructs with and without the element (Pendergrast and Hernandez, 1997; Sheldon *et al.*, 1993). Cells are first transfected with the wildtype LTR reporter construct and either a Tat expression plasmid or a mock plasmid. The cells are plated into microtitre plate wells in duplicate and the dilutions of the compounds DRB or TRB in 100% DMSO are added into the wells to maintain a constant 1% DMSO. Cells are
20 harvested and assayed for chloramphenicol acetyl transferase after 24 hours. General cytotoxicity assays are used to monitor cell death as described in Mancebo *et al.* Once the special sensitivity of the HIV-LTR is produced then the same assays using constructs in which the IST has been deleted are used. If the IST is what makes the LTR different from other promoters then the level of drug sensitivity decreases upon deletion of the element and the LTR
25 minus IST looks more like a normal promoter. These experiments are carried out in HeLa cells
30

that exhibit a low but reasonable level of Tat transactivation and in Jurkat cells that exhibit a much greater stimulation by Tat. If the IST element is the key for the uniqueness of the HIV-LTR then the FBI-1 protein characterized is a prime candidate for the factor responsible.

5 While the *in vivo* studies are underway the effect of the IST element is examined *in vitro*. Tat transactivation of the HIV-LTR is examined as in Example 6. First it is determined if the 50% inhibition point for the appearance of runoff from the LTR is different from that seen for other promoters such as the CMV promoter. If the difference is produced *in vitro* then LTR constructs are used with and without the IST element to probe for differences in the response to
10 the inhibitors of P-TEFb.

C. Development of *In vitro* Assays that Allow Precise Determination of the Effect of Tat.

The study of Tat transactivation *in vitro* has been hampered by an inability to obtain significant stimulation of the elongation in a pulse chase assay designed to separate initiation from elongation. It has been the rule rather than the exception to use unusual conditions to obtain significant stimulation of DRB sensitive transcripts from the HIV-LTR *in vitro*. In some labs a 30 minute “presynthesis” step was used (Marciniak and Sharp, 1991) and in others the inclusion of citrate (Parada and Roeder, 1996) was needed to see significant Tat transactivation. The inventor has found that a 20 minute continuous labeling assay in which the NTP concentrations were lowered to 50 μ M was sufficient to see a 12 fold stimulation.
15
20

The first step is to produce extracts from Jurkat cells, a B cell line that can support very high levels of Tat transactivation. The standard nuclear extract preparation described previously (Price *et al.*, 1987) that has yielded very active extracts containing all the factors necessary for
25 the efficient production of DRB-sensitive runoff transcripts is used. Cells are grown in suspension with high levels of fetal calf serum and are harvested in log growth. The extract is optimized for Tat transactivation with standard extract and template titrations as well as a condition search that includes titrations of mono- (KCl) and divalent (Mg^{++} and Zn^{++}) cations. The inventor has found that the source of Tat is important. A Jurkat cell line stably expressing

wildtype Tat and a Tat cys22gly mutation (Zauli *et al.*, 1995) is also useful. Extracts made from these cells should have native Tat that is active as long as the activation domain has no mutations. The negative control for *in vitro* transcription reactions is to use the mutant extract normalized with the CMV promoter or the HIV-LTR template lacking the TAR element.

5

If Tat transactivation is found to be robust in the newly derived system, assays are used that allow more insight into the step or steps affected by Tat. A pulse chase reaction is used but otherwise the conditions are kept similar to the continuous labeling conditions that normally work. Pulse reactions are long at first and then are shortened. Care is taken to determine nascent transcript length so that it can be determined if nascent transcripts need to contain TAR RNA or if generation of TAR RNA during the chase is sufficient. The LTR construct containing the lac repressor block at +200 described above is used to generate elongation complexes containing TAR RNA during the pulse and then IPTG is added at various times during the chase to observe the kinetic parameters of functional TAT/TAR interaction.

10

Data is consistent with Tat having an effect on initiation that is manifest as an effect on elongation during the continuous labeling assay . It is possible that Tat affects reinitiation of other polymerase molecules at promoters just used. This hypothesis is examined by using a multiple round assay that allows counting of the number of polymerases that initiate from a single promoter during the course of the reaction. Basically, the lac repressor is used to block polymerases that have entered productive elongation and have progressed about 700 bp into the template. The second polymerase that initiates from the same promoter is stopped about 30 bp earlier due to steric interference with the existing blocked polymerase (Szentirmay and Sawadogo, 1993). The length of the 30 bp ladder indicates the number of polymerases that reinitiate. IPTG allows all appropriately blocked polymerases to run off the end of the template at 800 bp. The reinitiation assay can be run under identical conditions to the continuous labeling conditions that have been shown to give reasonable Tat transactivation. If evidence of reinitiation being stimulated by Tat is seen then it is conclusive and dictates a path of research designed to probe the reinitiation. If there is no evidence of reinitiation, but Tat is seen to work

well, then reinitiation of transcription by Tat is ruled out as a major mechanism of Tat transactivation.

EXAMPLE 14
Tat Transactivation in Yeast

Both subunits of human P-TEFb are first transfected into yeast using a centromere plasmid with weak promoters driving the human cDNAs. The production of P-TEFb is followed with immunoblotting. It is first determined if human P-TEFb replaces the yeast CTD kinase, CTD K1, that has been recently demonstrated to be able to stimulate long transcripts in a HeLa extract system after inhibition of P-TEFb by DRB (Lee and Greenleaf, 1997; Sterner *et al.*, 1995; Lee and Greenleaf, 1992). Yeast lacking expression of the kinase subunit, CTK 1, are viable but have a cold sensitive phenotype and other growth defects. It is next determined if human P-TEFb reverses these effects in a CTK 1 null. If P-TEFb causes more normal growth, it is possible that CTD K1 is the functional equivalent of P-TEFb in yeast. Yeast expressing human P-TEFb is transfected with a construct containing the sequences from +1 to +80 of the HIV-LTR coupled to a standard yeast promoter CYC1. The hybrid promoter is used to drive a β -galactosidase reporter. The effect of Tat is determined by including a Tat expression cassette in the β -galactosidase reporter construct. The yeast system is then used to examine the interaction between human P-TEFb and Tat.

Alternatively, if Tat transactivation is not attainable, the yeast system may still be used in examining the interaction between Tat and P-TEFb if the biochemical approach suggests that the two proteins interact. A two hybrid system is set up with a Gal 4-Tat fusion protein as bait. The human P-TEFb expression plasmid is modified such that the kinase carries a knockout mutation and the cyclin subunit has an activation domain attached at the carboxyl-terminus. The reporter containing Gal 4 binding sites should be activated by the P-TEFb expression plasmid. If so, then deletions of the large subunit are carried out to see if the interaction assayed by reporter activity is lost. These studies are compared to studies that examine the ability of truncated versions of P-

TEFb to function in general elongation control *in vitro* and in Tat transactivation *in vitro* and *in vivo*.

* * *

5

All of the compositions and methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the compositions and methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

DRAFT - NOT FOR FILING

REFERENCES

The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated herein by reference.

- 5 Aboul-ela, F., Karn, J., and Varani, G. (1995). The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *Journal of Molecular Biology* 253, 313-332.
- 10 Akoulitchev, Makela, Weinberg, and Reinberg, "Requirement for TFIIH kinase activity in transcription by RNA polymerase II," *Nature*, 377:557-560, 1995.
- Allison, Wong, Fitzpatrick, Moyle, Ingles, *Molecular and Cellular Biology*, 8(1):321-329, 1988.
- In: *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988.
- Austin and Biggin, "In vitro transcription, Tfiih, Tfiie, Tfiiif, Tfidi," *Proceedings of the National Academy of Sciences of the United States of America*, 93:5788-5792, 1996.
- 15 Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1989). Current protocols in molecular biology (New York: Greene Publishing Associates and Wiley-Interscience).Bartolomei, Halden, Cullen, and Corden, *Molecular and Cellular Biology*, 8(1):330-339, 1988.
- Baskaran, Dahmus, Wang, *Proc. Natl. Acad. Sci. USA*, 90:11167-11171, 1993.
- 20 Bender, Thompson, Kuehl, "Differential expression of c-myb mRNA in murine B lymphomas by a block to transcription elongation," *Science*, 237:1473-1476, 1987.
- Bengal, Flores, Krauskopf, Reinberg, Aloni, "Role of the mammalian transcription factors IIF, IIS, and IIX during elongation by RNA polymerase II," *Mol. Cell. Biol.*, 11:1195-1206, 1991.
- 25 Bentley, "Regulation of transcriptional elongation by RNA polymerase II," *Curr. Opin. Genet. Dev.*, 5:210-216, 1995.
- Blau, Xiao, McCracken, O'Hare, Greenblatt, Bentley, "Three functional classes of transcriptional activation domains," *Molecular and Cellular Biology*, 16(5):2044-2055, 1996
- 30 Braddock, M., Thorburn, A.M., Kingsman, A.J., and Kingsman, S.M. (1991). Blocking of Tat-dependent HIV-1 RNA modification by an inhibitor of RNA polymerase II processivity. *Nature* 350, 439-441.
- Bradsher, Jackson, Conaway, and Conaway, "RNA polymerase II transcription factor SIII. I. Identification, purification, and properties," *J. Biol. Chem.*, 268:25587-25593, 1993a.

- Bradsher, Tan, McLaury, Conaway, Conaway, *J. Biol. Chem.*, 268:25594-25603, 1993b.
- Brake, D.A., Goudsmit, J., Krone, W.J., Schammel, P., Appleby, N., Meloen, R.H., and Debouck, C. (1990). Characterization of murine monoclonal antibodies to the tat protein from human immunodeficiency virus type 1. *Journal of Virology* 64, 962-965. Brickey and Greenleaf, *Genetics*, 140:599-613, 1995.
5
- Brent and Ptashne, *Cell*, 43:729-736, 1986
- Brutlag *et al.*, *CABIOS*, 6:237-245, 1990.
- Bullrich, MacLachlan, Sang, Druck, Veronese, Allen, Chiorazzi, Koff, Heubner, Croce *et al.*, "Chromosomal mapping of members of the cdc2 family of protein kinases, cdk3, cdk6, PISSLRE, and PITALRE, and a cdk inhibitor, p27Kip1, to regions involved in human cancer," *Cancer Research*, 55(6):1199-1205, 1995.
10
- Burton, Killeen, Sopta, Ortolan, Greenblatt, "RAP30/74: a general initiation factor that binds to RNA polymerase II," *Molecular and Cellular Biology*, 8(4):1602-1613, 1988.
- Cadena and Dahmus, *J. Biol. Chem.*, 262(26):12468-12474, 1987.
15
- Campbell, In: *Monoclonal Antibody Technology, Laboratory Techniques in Biochemistry and Molecular Biology*, Vol. 13, Burden and Von Knippenberg, Eds. pp. 75-83, Amsterdam, Elseview, 1984.
- Chambers, Wang, Burton, Dahmus, *J. Biol. Chem.*, 270:14962-14969, 1995.
20
- Chambers, R.S. and Kane, C.M. (1996). Purification and characterization of an RNA polymerase II phosphatase from yeast. *Journal of Biological Chemistry* 271, 24498-24504.
- Chen, Harless, Wright, and Kellem, "Identification and characterization of transcriptional arrest sites in exon 1 of the human adenosine deaminase gene," *Mol. Cell. Biol.*, 10:4555-4564, 1990.
25
- Chen, Innis, Sun, Wright, Kellem, "Sequence requirements for transcriptional arrest in exon 1 of the human adenosine deaminase gene," *Mol. Cell. Biol.*, 11:6248-6256, 1991.
- Chiang, P.W., Fogel, E., Jackson, C.L., Lieuallen, K., Lennon, G., Qu, X., Wang, S.Q., and Kurnit, D.M. (1996a). Isolation, sequencing, and mapping of the human homologue of the yeast transcription factor, SPT5. *Genomics* 38, 421-424.
30
- Chiang, P.W., Wang, S.Q., Smithivas, P., Song, W.J., Crombez, E., Akhtar, A., Im, R., Greenfield, J., Ramamoorthy, S., Van Keuren, M., Blackburn, C.C., Tsai, C.H., and Kurnit, D.M. (1996b). Isolation and characterization of the human and mouse homologues (SUPT4H and Supt4h) of the yeast SPT4 gene. *Genomics* 34, 368-375.

- Chien, Bartel, Sternglanz, Fields, "The two-hybrid system: A method to identify and clone genes for proteins that interact with a protein of interest," *Proc. Natl. Acad. Sci. USA*, 88:9578-9582, 1991.
- 5 Chinsky, Maa, Ramamurthy, Kellems, *J. Biol. Chem.*, 264:14561-14565, 1989.
- Chodosh, Fire, Samuels, Sharp, *J. Biol. Chem.*, 264:2250-2257, 1989.
- Chou and Fasman, "Conformational Parameters for Amino Acids in Helical, b-Sheet, and Random Coil Regions Calculated from Proteins," *Biochemistry*, 13(2):211-222, 1974b.
- 10 Chou and Fasman, "Empirical Predictions of Protein Conformation," *Ann. Rev. Biochem.*, 47:251-276, 1978b.
- Chou and Fasman, "Prediction of b-Turns," *Biophys. J.*, 26:367-384, 1979.
- Chou and Fasman, "Prediction of Protein Conformation," *Biochemistry*, 13(2):222-245, 1974a.
- Chou and Fasman, "Prediction of the Secondary Structure of Proteins from Their Amino Acid Sequence," *Adv. Enzymol. Relat. Areas Mol. Biol.*, 47:45-148, 1978a.
- 15 Chun, R.F. and Jeang, K.T. (1996). Requirements for RNA polymerase II carboxyl-terminal domain for transcription of human retroviruses, human T-cell lymphotropic virus, and HIV-1. *J. Biol. Chem.* 271, 27888-27894.
- Cisek and Corden, *Nature*, 339:679-684, 1989.
- Cismowski, Laff, Solomon, and Reed, *Mol. Cell. Biol.*, 15:2983-2992, 1995.
- 20 Collart, Tourkine, Belin, Vassalli, Jeanteur, Blanchard, "c-fos Gene transcription in murine macrophages is modulated by a calcium-dependent block to elongation in intron 1," *Mol. Cell. Biol.*, 11:2826-2831, 1991.
- Cujec, T.P., Cho, H., Maldonado, E., Meyer, J., Reinberg, D., and Peterlin, B.M. (1997). The human immunodeficiency virus transactivator Tat interacts with the RNA polymerase II holoenzyme. *Mol. Cell. Biol.* 17, 1817-1823.
- 25 Cullen, B.R. (1991). Regulation of HIV-1 gene expression. [Review]. *FASEB Journal* 5, 2361-2368.
- Dahmus, "Phosphorylation of the C-terminal domain of RNA polymerase II," *Biochim. Biophys. Acta Gene Struct. Expression*, 1261:171-182, 1995.

- Dahmus, "The role of multisite phosphorylation in the regulation of RNA polymerase II activity," [Review]. *Progress in Nucleic Acid Research and Molecular Biology*, 48:143-179, 1994.
- 5 Dahmus, M.E. (1996). Reversible phosphorylation of the c-terminal domain of RNA polymerase II. [review]. *J. Biol. Chem.* 271, 19009-19012.
- Dignam, Lebovitz, Roeder, "Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei," *Nucleic Acids Res.*, 11:1475-1489, 1983.
- 10 Dubois, Nguyen, Bellier, Bensaude, *J. Biol. Chem.*, 269:13331-13336, 1994.
- Durfee T *et al.*, "The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit," *Genes Dev.*, 7:555-569, 1993.
- 15 Dvir, Peterson, Knuth, Lu, Dynan, *Proc. Natl. Acad. Sci. USA*, 89:11920-11924, 1992.
- Egyhazi, Ossoinak, Pigon, Holmgren, Lee, Greenleaf, *Chromosoma*, 104:422-433, 1996.
- Feaver, Gileadi, Li, Kornberg, "CTD kinase associated with yeast RNA polymerase II initiation factor b," *Cell*, 67:1223-1230, 1991.
- 20 15 Feaver, Svejstrup, Henry, Kornberg, *Cell*, 79:1103-1109, 1994.
- Fetrow and Bryant, "New Programs for Protein Tertiary Structure Prediction," *Biotechnology*, 11:479-483, 1993.
- Fields and Song, "A novel genetic system to detect protein-protein interactions," *Nature*, 340:245-246, 1989.
- 25 20 Finkelstein, A., Kostrub, C.F., Li, J., Chavez, D.P., Wang, B.Q., Fang, S.M., Greenblatt, J., and Burton, Z.F. (1992). A cDNA encoding RAP74, a general initiation factor for transcription by RNA polymerase II. *Nature* 355, 464-467.
- Flores, Lu, Reinberg, *J. Biol. Chem.*, 267:2786-2793, 1992.
- 25 25 Flores, Maldonado, Reinberg, "Factors involved in specific transcription by mammalian RNA polymerase II. Factors IIE and IIF independently interact with RNA polymerase II," *J. Biol. Chem.*, 264:8913-8921, 1989.
- Fong, S.E., Smanik, P., Thais, T., Smith, M.C., and Jaskunas, S.R. (1997). Detection of specific human immunodeficiency virus type 1 tat-tar complexes in the presence of mild denaturing conditions. *Journal of Virological Methods* 66, 91-101.
- 30 30 Fraser, Sehgal, Darnell, *Proc. Natl. Acad. Sci. USA*, 76:2571-2575, 1979.

- Frohman, In: *PCR Protocols: A Guide to Methods and Applications*, Academic Press, N.Y., 1990.
- Garcia-Martinez, L.F., Ivanov, D., and Gaynor, R.B. (1997a). Association of Tat with purified HIV-1 and HIV-2 transcription preinitiation complexes. *J. Biol. Chem.* 272, 6951-6958.
- 5 Garcia-Martinez, L.F., Mavankal, G., Neveu, J.M., Lane, W.S., Ivanov, D., and Gaynor, R.B. (1997b). Purification of a Tat-associated kinase reveals a TFIIH complex that modulates HIV-1 transcription. *EMBO J.* 16, 2836-2850.
- Garcia and Gaynor, "The human immunodeficiency virus type-1 long terminal repeat and its role in gene expression," [Review]. *Progress in Nucleic Acid Research and Molecular Biology*, 49:157-196, 1994.
- 10 Garriga, J., Mayol, X., and Grana, X. (1996a). The CDC2-related kinase PITALRE is the catalytic subunit of active multimeric protein complexes. *Biochem. J.* 319, 293-298.
- Gefter et al., *Somatic Cell Genet.*, 3:231-236, 1977.
- 15 Goding, In: *Monoclonal Antibodies: Principles and Practice*, 2d ed., Orlando, Fla., Academic Press, 1986, pp. 60-61, 65-66, 71-74.
- Gossen and Bujard, *Proc. Natl. Acad. Sci. USA*, 89:5547-5551, 1992.
- Gossen et al., *Science*, 268:1766-1769, 1995.
- 20 Grana, De Luca, Sang, Fu, Claudio, Rosenblatt, Morgan, Giordano, "PITALRE, a nuclear CDC2-related protein kinase that phosphorylates the retinoblastoma protein *in vitro*", *Proc. Natl. Acad. Sci. USA*, 91:3834-3838, 1994.
- Guo and Price, "Mechanism of DmS-II-mediated pause suppression by *Drosophila* RNA polymerase II," *J. Biol. Chem.*, 268:18762-18770, 1993.
- Hair and Morgan, *Mol. Cell. Biol.*, 13:7925-7934, 1993.
- 25 Harper, Adami, Wei, Keyomarsi, Elledge, "The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases," *Cell*, 75:805-816, 1993.
- Herrmann and Rice, "Lentivirus Tat proteins specifically associate with a cellular protein kinase, TAK, that hyperphosphorylates the carboxyl-terminal domain of the large subunit of RNA polymerase II: Candidate for a Tat cofactor," *J. Virol.*, 69:1612-1620, 1995.
- 30 Herrmann and Rice, "Specific interaction of the human immunodeficiency virus Tat proteins with a cellular protein kinase," *Virology*, 197:601-608, 1993.

- Herrmann, Gold, Rice, "Viral transactivators specifically target distinct cellular protein kinases that phosphorylate the RNA polymerase II C-terminal domain," *Nucleic Acids Research*, 24(3):501-508, 1996.
- Hopp, U.S. Patent 4,554,101.
- 5 Irie, K., Nomoto, S., Miyajima, I., and Matsumoto, K. (1991). SGV1 encodes a CDC28/cdc2-related kinase required for a G alpha subunit-mediated adaptive response to pheromone in *S. cerevisiae*. *Cell* 65, 785-795.
- Izban and Luse, "SII-facilitated transcript cleavage in RNA polymerase II complexes stalled early after initiation occurs in primarily dinucleotide increments," *J. Biol. Chem.*, 10 268:12864-12873, 1993.
- Jameson and Wolf, "The Antigenic Index: A Novel Algorithm for Predicting Antigenic Determinants," *Comput. Appl. Biosci.*, 4(1):181-186, 1988.
- Johnston, *Microbiol. Rev.* 51:458-476, 1987).
- 15 Jones and Peterlin, "Control of RNA initiation and elongation at the HIV-1 promoter," [Review]. *Annual Review of Biochemistry*, 63:717-743, 1994.
- Kang and Dahmus, "RNA polymerases IIA and IIO have distinct roles during transcription from the TATA-less murine dihydrofolate reductase promoter," *J. Biol. Chem.*, 268:25033-25040, 1993.
- Karin and Hunter, *Curr. Biol.*, 5:747-757, 1995.
- 20 Kash, Innis, Jackson, Kellems, "Functional analysis of a stable transcription arrest site in the first intron of the murine adenosine deaminase gene," *Mol. Cell. Biol.*, 13:2718-2729, 1993.
- Keegan, Gill, Ptashne, *Science*, 231:699-704, 1986.
- 25 Keen, N.J., Gait, M.J., and Karn, J. (1996). Human immunodeficiency virus type-1 Tat is an integral component of the activated transcription-elongation complex. *Proc. Natl. Acad. Sci. USA* 93, 2505-2510.
- Kephart, Marshall, Price, "Stability of *Drosophila* RNA polymerase II elongation complexes *in vitro*," *Mol. Cell. Biol.*, 12:2067-2077, 1992.
- Kephart, Wang, Burton, Price, "Functional analysis of *Drosophila* factor 5 (TFIIF), a general transcription factor," *J. Biol. Chem.*, 269:13536-13543, 1994.
- 30 Kerppola and Kane, *FASEB J.*, 5:2833-2842, 1991.
- Kessler, Ben-Asher, Aloni, "Elements modulating the block of transcription elongation at the adenovirus 2 attenuation site," *J. Biol. Chem.*, 264:9785-9790, 1989.

- Kessler, Ben-Asher, Resnekov, Hatini, Bengal, Aloni, "A 21-base pair DNA fragment directs transcription attenuation within the simian virus 40 late leader," *J. Biol. Chem.*, 266:13019-13027, 1991.
- Kim and Dahmus, *J. Biol. Chem.*, 264:3169-3176, 1989.
- 5 Kohler and Milstein, *Eur. J. Immunol.*, 6:511-519, 1976.
- Kohler and Milstein, *Nature*, 256:495-497, 1975.
- Koleske and Young, *Trends Biochem. Sci.*, 20:113-116, 1995.
- Kozak, *Nucleic Acids Res.*, 15:8125-8148, 1987.
- 10 Krauskopf, Ben-Asher, Aloni, "Minute virus of mice infection modifies cellular transcription elongation," *J. Virol.*, 68:2741-2745, 1994.
- Krauskopf, Bengal, Aloni, *Mol. Cell. Biol.*, 11:3515-3521, 1991.
- Krauskopf, Resnekov, Aloni, "A *cis* downstream element participates in regulation of *in vitro* transcription initiation from the P38 promoter of minute virus of mice," *J. Virol.*, 64:354-360, 1990.
- 15 Krumm, Hickey, and Groudine, "Promoter-proximal pausing of RNA polymerase II defines a general rate-limiting step after transcription initiation," *Genes and Development*, 9:559-572, 1995.
- Krumm, Meulia, Brunvand, Groudine, *Genes Dev.*, 6:2201-2213, 1992.
- 20 Kyte, Doolittle, "A simple method for displaying the hydropathic character of a protein," *J. Mol. Biol.*, 157(1):105-132, 1982.
- Laspia, Rice, Mathews, "HIV-1 Tat protein increases transcriptional initiation and stabilizes elongation," *Cell*, 59:283-292, 1989.
- Laub, Jakobovits, Aloni, *Proceedings of the National Academy of Sciences of the United States of America*, 77:3297-3301, 1980.
- 25 Lee and Greenleaf, "CTD kinase large subunit is encoded by CTK1, a gene required for normal growth of *Saccharomyces cerevisiae*," *Gene Expression*, 1:149-167, 1992.
- Lee and Greenleaf, *Proc. Natl. Acad. Sci. USA*, 86:3624-3628, 1989.
- Lee, J.M. and Greenleaf, A.L. (1997). Modulation of RNA polymerase II elongation efficiency by c-terminal heptapeptide repeat domain kinase I. *J. Biol. Chem.* 272, 10990-10993.
- 30 Li and Kornberg, *Proc. Natl. Acad. Sci. USA*, 91:2362-2366, 1994.

- Liao, Zhang, Jeffery, Koleske, Thompson, Chao, Viljoen, Van Vuuren, Young, *Nature*, 374:193-196, 1995.
- Lu, Zawel, Fisher, Egly, Reinberg, "Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II," *Nature*, 358:641-645, 1992.
- 5 MacLachlan, T.K., Sang, N., and Giordano, A. (1995). Cyclins, cyclin-dependent kinases and cdk inhibitors: implications in cell cycle control and cancer. [Review] [235 refs]. Critical Reviews in Eukaryotic Gene Expression 5, 127-156.
- 10 Mäkelä, Parvin, Kim, Huber, Sharp, Weinberg, "A kinase-deficient transcription factor TFIIH is functional in basal and activated transcription," *Proc. Natl. Acad. Sci. USA*, 92:5174-5178, 1995.
- Maldonado and Reinberg, "News on initiation and elongation of transcription by RNA polymerase II," *Curr. Opin. Cell Biol.*, 7:352-361, 1995.
- Marciniak and Sharp, "HIV-1 Tat protein promotes formation of more-processive elongation complexes," *EMBO J.*, 10:4189-4196, 1991.
- 15 Marshall and Price, "Control of formation of two distinct classes of RNA polymerase II elongation complexes," *Mol. Cell. Biol.*, 12:2078-2090, 1992.
- Marshall and Price, "Purification of P-TEFb, a transcription factor required for the transition into productive elongation," *J. Biol. Chem.*, 270:12335-12338, 1995.
- 20 Marshall, N.F., Peng, J.M., Xie, Z., and Price, D.H. (1996). Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J. Biol. Chem.* 271, 27176-27183.
- Mavankal, G., Ou, S.H.I., Oliver, H., Sigman, D., Gaynor, R.B., Transcription factors, and Gene expression (1996). Human immunodeficiency virus type 1 and 2 Tat proteins specifically interact with RNA polymerase II. *Proc. Natl. Acad. Sci. USA* 93, 2089-2094.
- 25 Mechti, Piechaczyk, Blanchard, Jeanteur, Lebleu, "Sequence requirements for premature transcription arrest within the first intron of the mouse c-fos gene," *Mol. Cell. Biol.*, 11:2832-2841, 1991.
- Meulia, Krumm, Groudine, "Distinct properties of c-myc transcriptional elongation are revealed in *Xenopus* oocytes and mammalian cells and by template titration, 5,6-dichloro-1-b-D-ribofuranosylbenzimidazole (DRB), and promoter mutagenesis," *Mol. Cell. Biol.*, 13:5647-5658, 1993.
- 30 Middleton and Morgan, *Mol. Cell Biol.*, 10:727-735, 1990.

- DRAFT EDITION
- Miller, Asselin, Dufort, Yang, Gupta, Marcu, Nepveu, "A cis-acting element in the promoter region of the murine c-myc gene is necessary for transcriptional block," *Mol. Cell. Biol.*, 9:5340-5349, 1989.
- Morgan, Principles of CDK regulation, *Nature* 374:131-134, 1995.
- 5 Nakamura *et al.*, In: *Enzyme Immunoassays: Heterogeneous and Homogeneous Systems*, Chapter 27, 1987.
- Neenhold, H.R. and Rana, T.M. (1995). Major groove opening at the HIV-1 Tat binding site of TAR RNA evidenced by a rhodium probe. *Biochemistry* 34, 6303-6309.
- Nonet, Sweetser, Young, *Cell*, 50:909-915, 1987.
- 10 O'Brien, Hardin, Greenleaf, Lis, *Nature*, 370:75-77, 1994.
- Parada, C.A. and Roeder, R.G. (1996). Enhanced processivity of RNA polymerase II triggered by Tat-induced phosphorylation of its carboxy-terminal domain. *Nature* 384, 375-378.
- 15 Payne and Dahmus, "Partial purification and characterization of two distinct protein kinases that differentially phosphorylate the carboxyl-terminal domain of RNA polymerase subunit IIa," *J. Biol. Chem.*, 268:80-87, 1993.
- Pendergrast, P.S. and Hernandez, N. (1997). RNA-targeted activators, but not DNA-targeted activators, repress the synthesis of short transcripts at the human immunodeficiency virus type 1 long terminal repeat. *J. Virol.* 71, 910-917.
- 20 Pessler, F., Pendergrast, P.S., and Hernandez, N. (1997). Purification and characterization of FBI-1, a cellular factor that binds to the human immunodeficiency virus type 1 inducer of short transcripts. *Molecular & Cellular Biology* 17, 3786-3798.
- Peterson, Dvir, Anderson, Dynan, "DNA binding provides a signal for phosphorylation of the RNA polymerase II heptapeptide repeats," *Genes Dev.*, 6:426-438, 1992.
- 25 Price, Sluder, Greenleaf, "Dynamic interaction between a Drosophila transcription factor and RNA polymerase II," *Mol. Cell Biol.*, 9:1465-1475, 1989.
- Price, Sluder, Greenleaf, "Fractionation of transcription factors for RNA polymerase II from *Drosophila K_c* cell nuclear extracts," *J. Biol. Chem.*, 262:3244-3255, 1987.
- Ramamurthy, Maa, Harless, Wright, Kellems, *Mol. Cell. Biol.*, 10:1484-1491, 1990.
- Rasmussen and Lis, *J. Mol. Biol.*, 252:522-535, 1995.
- 30 Reddy and Reddy, "Differential binding of nuclear factors to the intron 1 sequences containing the transcriptional pause site correlates with c-myb expression," *Proc. Natl. Acad. Sci. USA*, 86:7326-7330, 1989.

- P&G
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
- Reinberg and Roeder, "Factors involved in specific transcription by mammalian RNA polymerase II. Transcription factor IIS stimulates elongation of RNA chains," *J. Biol. Chem.*, 262:3331-3337, 1987.
- 5 Reines, "Elongation factor-dependent transcript shortening by template-engaged RNA polymerase II," *J. Biol. Chem.*, 267:3795-3800, 1992.
- Remington's Pharmaceutical Sciences*, 15th Edition, 1035-1038 and 1570-1580.
- Resnekov and Aloni, *Proc. Natl. Acad. Sci. USA*, 86:12-16, 1989.
- Resnekov, Ben-Asher, Bengal, Choder, Hay, Kessler, Ragimov, Seiberg, Skolnik-David, and Aloni, "Transcription termination in animal viruses and cells," *Gene*, 72:91-104, 1988.
- 10 Resnitzky *et al.*, *Mol. Cell. Biol.*, 14:1669-1679, 1994.
- Rhim, H. and Rice, A.P. (1994). Functional significance of the dinucleotide bulge in stem-loop1 and stem-loop2 of HIV-2 TAR RNA. *Virology* 202, 202-211.
- Rice, A.P. and Carlotti, F. (1990). Structural analysis of wild-type and mutant human immunodeficiency virus type 1 Tat proteins. *J. Virol.* 64, 6018-6026.
- 15 Roberts and Bentley, *EMBO J.*, 11:1085-1093, 1992.
- Rougvie and Lis, "Postinitiation transcriptional control in *Drosophila melanogaster*," *Mol. Cell. Biol.*, 10:6041-6045, 1990.
- Rougvie and Lis, "The RNA Polymerase II Molecule at the 5' End of the Uninduced hsp70 Gene of D. Melanogaster is Transcriptionally Engaged," *Cell*, 54:795-804, 1988.
- 20 Roy, Adamczewski, Seroz, Vermeulen, Tassan, Schaeffer, Nigg, Hoeijmakers, and Egly, *Cell*, 79:1093-1101, 1994.
- Sambrook J, Fritsch EF, Maniatis, T, In: *Molecular cloning: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, New York, 1989.
- 25 Sanchez-Pescador, Power, Barr, Steimer, Stempien, Brown-Shimer, Gee, Renard, Randolph, Levy *et al.*, "Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2)," *Science*, 227(4686):484-492, 1985.
- Schneider-Schaulies, Schimpl, Wecker, *Eur. J. Immunol.*, 17:713-718, 1987.
- Sehgal, Darnell, Tamm, *Cell*, 9:473-480, 1976.
- 30 Serizawa, Conaway, Conaway, "A carboxyl-terminal-domain kinase associated with RNA polymerase II transcription factor d from rat liver," *Proc. Natl. Acad. Sci. USA*, 89:7476-7480, 1992.

- RECEIVED
2001
15
- Serizawa, Mäkelä, Conaway, Conaway, Weinberg, Young, *Nature*, 374:280-282, 1995.
- Sheldon, M., Ratnasabapathy, R., and Hernandez, N. (1993). Characterization of the inducer of short transcripts, a human immunodeficiency virus type 1 transcriptional element that activates the synthesis of short RNAs. *Mol. Cell. Biol.* 13, 1251-1263.
- 5 Shiekhattar, Mermelstein, Fisher, Drapkin, Dynlacht, Wessling, Morgan, and Reinberg, *Nature*, 374:283-287, 1995.
- Shilatifard, Lane, Jackson, Conaway, Conaway, "An RNA polymerase II elongation factor encoded by the human ELL gene," *Science*, 271:1873-1876, 1996.
- 10 SivaRaman, Reines, Kane, "Purified elongation factor SII is sufficient to promote read-through by purified RNA polymerase II at specific termination sites in the human histone H3.3 gene," *J. Biol. Chem.*, 265:14554-14560, 1990.
- Sluder, Greenleaf, Price, "Properties of a Drosophila RNA polymerase II elongation factor," *J. Biol. Chem.*, 264:8963-8969, 1989.
- 15 Spencer and Groudine, "Molecular analysis of the c-myc transcription elongation block: Implications for the generation of Burkitt's lymphoma," *Ann. NY Acad. Sci.*, 599:12-28, 1990a.
- Spencer and Groudine, *Oncogene*, 5:777-786, 1990b.
- 20 Stern, Lee, Hardin, Greenleaf, "The yeast carboxyl-terminal repeat domain kinase CTDK-I is a divergent cyclin-cyclin-dependent kinase complex," *Molecular and Cellular Biology*, 15:5716-5724, 1995.
- Stone and Reinberg, *J. Biol. Chem.*, 267:6353-6360, 1992.
- Strobl and Eick, "Hold back of RNA polymerase II at the transcription start site mediates down-regulation of c-myc *in vivo*," *EMBO J.*, 11:3307-3314, 1992.
- 25 Szentirmay, M.N. and Sawadogo, M. (1993). Synthesis of reinitiated transcripts by mammalian RNA polymerase II is controlled by elongation factor SII. *EMBO J.* 12, 4677-4684.
- Tamm and Kikuchi, *Proceedings of the National Academy of Sciences of the United States of America*, 76:5750-5754, 1979.
- Tamm, Kikuchi, Darnell, Salditt-Georgieff, *Biochemistry*, 19:2743-2748, 1980.
- 30 *Tet Expression Systems and Cell Lines* (July 1996), CLONTECHniques XI(3):2-5, 1996.
- Thompson, Steinberg, Aronson, Burgess, *J. Biol. Chem.*, 264:11511-11520, 1989.
- Valay, Simon, Dubois, Bensaude, Facca, Faye, *J. Mol. Biol.*, 249:535-544, 1995.

- Venetianer, Dubois, Nguyen, Bellier, Seo, Bensaude, *Eur. J. Biochem.*, 233:83-92, 1995.
- Wampler and Kadonaga, "Functional analysis of *Drosophila* transcription factor IIB," *Genes Dev.*, 6:1542-1552, 1992.
- Wampler, Tyree, Kadonaga, "Fractionation of the general RNA polymerase II transcription factors from *Drosophila* embryos," *J. Biol. Chem.*, 265:21223-21231, 1990.
- Weinberger, C. et al., *Science*, 228:740-742, 1985.
- Wolf et al., "An Integrated Family of Amino Acid Sequence Analysis Programs," *Comput. Appl. Biosci.*, 4(1):187-191, 1988.
- Wright and Bishop, "DNA sequences that mediate attenuation of transcription from the mouse protooncogene myc," *Proc. Natl. Acad. Sci. USA*, 86:505-509, 1989.
- Wright, "Regulation of eukaryotic gene expression by transcriptional attenuation," *Molecular Biology of the Cell*, 4:661-668, 1993.
- Xie and Price, "Purification of an RNA polymerase II transcript release factor from *Drosophila*," *J. Biol. Chem.*, 271:11043-11046, 1996.
- Yang, Herrmann, Price, "The human immunodeficiency virus Tat proteins specifically associate with TAK *in vivo* and require the carboxy-terminal domain of RNA polymerase II for function," *Journal of Virology*, 70:4576-4584, 1996.
- Yankulov, K., Blau, J., Purton, T., Roberts, S., Bentley, "Transcriptional elongation by RNA polymerase II is stimulated by transactivators," *Cell*, 77:749-759, 1994.
- Yankulov, Pandes, McCracken, Bouchard, Bentley, "TFIIP functions in regulating transcriptional elongation by RNA polymerase II in *Xenopus* oocytes," *Molecular and Cellular Biology*, 16:3291-3299, 1996.
- Yankulov, Yamashita, Roy, Egly, Bentley, "The transcriptional elongation inhibitor 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole inhibits transcription factor IIH-associated protein kinase," *J. Biol. Chem.*, 270:23922-23925, 1995.
- Yue, Favot, Dunn, Cassady, Hume, "Expression of mRNA encoding the macrophage colony-stimulating factor receptor (*c-fms*) is controlled by a constitutive promoter and tissue-specific transcription elongation," *Mol. Cell. Biol.*, 13:3191-3201, 1993.
- Zandomeni, Zandomeni, Shugar, Weinmann, *J. Biol. Chem.*, 261:3414-3419, 1986.
- Zauli, G., Gibellini, D., Caputo, A., Bassini, A., Negrini, M., Monne, M., Mazzoni, M., and Capitani, S. (1995). The human immunodeficiency virus type-1 Tat protein upregulates Bcl-2 gene expression in Jurkat T-cell lines and primary peripheral blood mononuclear cells. *Blood* 86, 3823-3834.

Zehring, Lee, Weeks, Jokerst, and Greenleaf, *Proc. Natl. Acad. Sci. USA*, 85:3698-3702, 1988.

Zhou and Sharp, *EMBO Journal*, 14:321-328, 1995.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Price, David H.
- (ii) TITLE OF INVENTION: P-TEFb COMPOSITIONS, METHODS AND SCREENING ASSAYS
- (iii) NUMBER OF SEQUENCES: 68
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Arnold, White & Durkee
 - (B) STREET: P.O. Box 4433
 - (C) CITY: Houston
 - (D) STATE: TX
 - (E) COUNTRY: USA
 - (F) ZIP: 77210-4433
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: US Unknown
 - (B) FILING DATE: Concurrently Herewith
 - (C) CLASSIFICATION: Unknown
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Fussey, Shelley P.M.
 - (B) REGISTRATION NUMBER: 39,458
 - (C) REFERENCE/DOCKET NUMBER: IOWA:012
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (512) 418-3000
 - (B) TELEFAX: (512) 418-3131

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1457 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 115..1326

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TGTTGAGTCA ACAGCTGTAG ATACACCAAT TGTTGCCGAT TTCTTCCTT TCGACTGTCG	60
GCTTCTCGCG AAACGTGAT TGTGAAAATT GTACAAATAG AGGCAAATTT AACC ATG Met 1	117
GCG CAC ATG TCC CAC ATG CTC CAG CAG CCT TCG GGG TCG ACG CCC TCC Ala His Met Ser His Met Leu Gln Gln Pro Ser Gly Ser Thr Pro Ser 5 10 15	165
AAC GTG GGC TCC AGC TCA TCG CGC ACG ATG TCC CTG ATG GAG AAA CAA Asn Val Gly Ser Ser Ser Arg Thr Met Ser Leu Met Glu Lys Gln 20 25 30	213
AAG TAC ATC GAG GAC TAC GAC TTT CCC TAC TGC GAC GAG AGC AAC AAA Lys Tyr Ile Glu Asp Tyr Asp Phe Pro Tyr Cys Asp Glu Ser Asn Lys 35 40 45	261
TAC GAA AAG GTG GCG AAA ATT GGC CAA GGC ACC TTC GGA GAG GTT TTT Tyr Glu Lys Val Ala Lys Ile Gly Gln Gly Thr Phe Gly Glu Val Phe 50 55 60 65	309
AAG GCT CGC GAG AAA AAG GGC AAC AAG AAG TTT GTG GCC ATG AAG AAG Lys Ala Arg Glu Lys Lys Gly Asn Lys Phe Val Ala Met Lys Lys 70 75 80	357
GTG CTG ATG GAC AAC GAA AAG GAG GGC TTT CCC ATC ACG GCT CTG CGA Val Leu Met Asp Asn Glu Lys Glu Gly Phe Pro Ile Thr Ala Leu Arg 85 90 95	405
GAG ATC CGC ATC CTG CAG CTG CTA AAG CAC GAG AAC GTG GTG AAT CTG Glu Ile Arg Ile Leu Gln Leu Lys His Glu Asn Val Val Asn Leu 100 105 110	453
ATC GAG ATC TGC CGC ACC AAG GCC ACC GCC ACG AAT GGT TAC AGA TCC Ile Glu Ile Cys Arg Thr Lys Ala Thr Ala Thr Asn Gly Tyr Arg Ser 115 120 125	501
ACC TTC TAT TTG GTC TTT GAT TTC TGC GAA CAC GAT TTG GCA GGT CTT Thr Phe Tyr Leu Val Phe Asp Phe Cys Glu His Asp Leu Ala Gly Leu 130 135 140 145	549
CTG TCC AAC ATG AAC GTC AAG TTC AGT CTG GGC GAG ATT AAG AAG GTT Leu Ser Asn Met Asn Val Lys Phe Ser Leu Gly Glu Ile Lys Lys Val 150 155 160	597
ATG CAG CAG CTT TTA AAC GGT TTG TAT TAC ATC CAC AGC AAC AAG ATC Met Gln Gln Leu Leu Asn Gly Leu Tyr Tyr Ile His Ser Asn Lys Ile 165 170 175	645

CTG CAC CGA GAC ATG AAA GCT GCC AAC GTG CTG ATT ACC AAG CAT GGC Leu His Arg Asp Met Lys Ala Ala Asn Val Leu Ile Thr Lys His Gly 180 185 190	693
ATC TTA AAG CTG GCT GAC TTT GGC TTG GCC CGT GCT TTT AGC ATT CCA Ile Leu Lys Leu Ala Asp Phe Gly Leu Ala Arg Ala Phe Ser Ile Pro 195 200 205	741
AAG AAC GAG AGT AAG AAT CGC TAT ACC AAT CGC GTA GTA ACC TTG TGG Lys Asn Glu Ser Lys Asn Arg Tyr Thr Asn Arg Val Val Thr Leu Trp 210 215 220 225	789
TAC CGG CCG CCT GAG CTG CTA CTT GGT GAC CGC AAC TAT GGT CCA CCC Tyr Arg Pro Pro Glu Leu Leu Leu Gly Asp Arg Asn Tyr Gly Pro Pro 230 235 240	837
G TG GAC ATG TGG GGA GCC GGC TGC ATA ATG GCC GAG ATG TGG ACA CGC Val Asp Met Trp Gly Ala Gly Cys Ile Met Ala Glu Met Trp Thr Arg 245 250 255	885
TCG CCC ATC ATG CAA GGC AAT ACG GAG CAG CAG CAG TTA ACC TTT ATT Ser Pro Ile Met Gln Gly Asn Thr Glu Gln Gln Leu Thr Phe Ile 260 265 270	933
TCG CAG CTA TGC GGC TCC TTT ACG CCG GAC GTG TGG CCG GGA GTG GAG Ser Gln Leu Cys Gly Ser Phe Thr Pro Asp Val Trp Pro Gly Val Glu 275 280 285	981
GAG CTG GAG CTG TAC AAA TCC ATC GAG CTG CCA AAG AAC CAG AAG CGT Glu Leu Glu Leu Tyr Lys Ser Ile Glu Leu Pro Lys Asn Gln Lys Arg 290 295 300 305	1029
CGA GTC AAG GAG CGC CTG CGT CCG TAT GTC AAG GAT CAA ACC GGC TGT Arg Val Lys Glu Arg Leu Arg Pro Tyr Val Lys Asp Gln Thr Gly Cys 310 315 320	1077
GAT CTA TTG GAC AAA TTG CTG ACC CTT GAT CCC AAG AAA CGC ATC GAT Asp Leu Leu Asp Lys Leu Leu Thr Leu Asp Pro Lys Arg Ile Asp 325 330 335	1125
GCG GAC ACA GCT CTG AAT CAC GAC TTC TTC TGG ACG GAT CCC ATG CCC Ala Asp Thr Ala Leu Asn His Asp Phe Phe Trp Thr Asp Pro Met Pro 340 345 350	1173
AGC GAC TTG AGC AAG ATG CTG TCC CAG CAC CTG CAG AGC ATG TTC GAG Ser Asp Leu Ser Lys Met Leu Ser Gln His Leu Gln Ser Met Phe Glu 355 360 365	1221
TAC CTG GCG CAG CCA CGC CGC AGC AAC CAG ATG CGC AAC TAT CAC CAG Tyr Leu Ala Gln Pro Arg Arg Ser Asn Gln Met Arg Asn Tyr His Gln 370 375 380 385	1269

CAA CTG ACC ACC ATG AAC CAG AAG CCC CAG GAC AAC AGT ATG ATT GAC Gln Leu Thr Thr Met Asn Gln Lys Pro Gln Asp Asn Ser Met Ile Asp	390	395	400	1317
CGG GTT TGG TAGACTGCCA GAGGTGTACG CACCCGACTA ATAGTTCTC Arg Val Trp				1366
ACCTTCAACT AGCGTTAGGT TATTAGTTA GTGTACAATA AAAATATTGG CATTGCATT				1426
AGCGCTTGCT CCAAATATAA AAAAAAAA A				1457

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 404 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ala His Met Ser His Met Leu Gln Gln Pro Ser Gly Ser Thr Pro				
1	5	10	15	
Ser Asn Val Gly Ser Ser Ser Arg Thr Met Ser Leu Met Glu Lys				
20	25	30		
Gln Lys Tyr Ile Glu Asp Tyr Asp Phe Pro Tyr Cys Asp Glu Ser Asn				
35	40	45		
Lys Tyr Glu Lys Val Ala Lys Ile Gly Gln Gly Thr Phe Gly Glu Val				
50	55	60		
Phe Lys Ala Arg Glu Lys Lys Gly Asn Lys Lys Phe Val Ala Met Lys				
65	70	75	80	
Lys Val Leu Met Asp Asn Glu Lys Glu Gly Phe Pro Ile Thr Ala Leu				
85	90	95		
Arg Glu Ile Arg Ile Leu Gln Leu Leu Lys His Glu Asn Val Val Asn				
100	105	110		
Leu Ile Glu Ile Cys Arg Thr Lys Ala Thr Ala Thr Asn Gly Tyr Arg				
115	120	125		
Ser Thr Phe Tyr Leu Val Phe Asp Phe Cys Glu His Asp Leu Ala Gly				
130	135	140		
Leu Leu Ser Asn Met Asn Val Lys Phe Ser Leu Gly Glu Ile Lys Lys				
145	150	155	160	

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4328 base pairs
 - (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CAGCCCTGCC	GACGGCCATA	CTTGAAAATA	CATTTTTTC	TGCAAAGTTT	GTCATTGTCA	60
CTGTGTGAAT	GGAATCTGTG	ATGTGTTGTG	GAATTAaaaaaa	CGTCAAGTAA	ACAACCCGTA	120
ATGGTTAAAG	TGCACGGCGA	AAGCAGTGC	AATAACTATG	AATTGATACA	AAAGTTGCAT	180
AACACGTGCG	CTGGTGTGCG	GGTTAGTGTG	TTTTTCTGTCT	CGTTTCGTTT	CCGCCGCGAGT	240
CGCAGTTTCC	AAAAAAACCTC	ACCACACCAT	ACCATCTCCA	CCACGCACAC	ACACACACAA	300
ACAAACACGC	AGAGACGCGG	CGGCGGAAAAA	AGTGTGCGGA	CCGCGGATT	AACCCCTCGT	360
TCCAAACCCA	AATTGGAGTC	TCCCaaaaaac	AGCGAAATAT	CGAGTGTGGC	TTAGCCGATG	420
TGCCGTGCGA	TCCCCACTGC	CCCTTCCGTA	CCGCTGCCAC	CCCCGCCACA	GCAGCAACGC	480
ACACGGATAC	GGACACAGAC	ACCAATACCA	GCGCACTCAA	GCACGGCCGA	CAAAGAAAGA	540
GCGCTCTCCC	TTCCCTTTG	TACAGTTAGT	TCCTACAGCT	GAATCAGCCA	AAAGAAATTA	600
CTAGGTCCAT	TCCGAGGCGC	AGTTTGCATG	TGAAACGGAG	GTCCCCGCAT	AACCACGCGG	660
AACCCGAAAT	TCCAGATCCC	CATCTCCGCT	GCACGGATAA	AGGAAACATA	CAACCATGAG	720
TCTCCTAGCC	ACGCCAATGC	CCCAGGGGGC	CACCGCCTCA	TCTTCTTCAT	CCGCCTCCGC	780
GGCCGCCTCG	GCCAGCGGGA	TTCCAATCAC	CGCCAACAAC	AACCTGCC	TCGAGAAGGA	840
CAAGATCTGG	TACTTCAGCA	ACGATCAGCT	GGCCAATTG	CCAAGCAGAA	GATGCGGCAT	900
CAAGGGCGAC	GATGAGCTGC	AGTACCGCCA	GATGACCGCC	TATCTGATAC	AGGAAATGGG	960
TCAGCGTCTG	CAGGTGTCCC	AACTGTGCAT	CAACACGGCC	ATTGTGTACA	TGCATCGGTT	1020
CTACGCCTTT	CACTCCTTCA	CCCACTTCA	TCGCAACTCC	ATGGCGTCGG	CGAGCCTCTT	1080
CTTGGCCGCC	AAGGTAGAAG	AGCAACCGCG	GAAGCTGGAG	CATGTTATTC	GGGCCGCCAA	1140
CAAAGTGCCTG	CCGCCGACCA	CCGAGCAGAA	TTACGCCAA	CTCGCCAGG	AGCTTGTGTT	1200
CAACGAGAAC	GTGCTCCTGC	AGACGCTGGG	CTTCGATGTG	GCCATCGATC	ATCCGCACAC	1260
GCATGTGGTG	CGCACCTGCC	AGCTGGTCAA	AGCATGCAAG	GATCTGGCGC	AGACATCGTA	1320
CTTCTTGGCC	TCGAACAGCC	TGCATCTGAC	CTCGATGTG	CTCCAATATC	GCCCCACGGT	1380
CGTAGCCTGT	TTCTGCATTT	ACCTAGCCTG	CAAGTGGTCC	CGATGGGAGA	TCCCCCAGTC	1440
GACCGAGGGC	AAGCACTGGT	TCTACTATGT	GGACAAGACG	GTCTCGCTGG	ATTTGCTAAA	1500

GCAGCTGACA	GATGAGTTCA	TCGCTATCTA	TGAGAACAGGC	CCGGCCCCGTC	TGAAGTCTAA	1560
GCTTAACTCG	ATCAAGGCAGA	TCGCCAGGG	AGCCAGCAAT	CGGACAGCTA	ACAGCAAGGA	1620
CAAACCAAAG	GAGGACTGGA	AGATCACCGA	GATGATGAAG	GGCTACCACT	CAAACATCAC	1680
GACACCACCA	GAGCTGTTAA	ACGGCAACGA	CAGCCGGAT	CGGGACCGAG	ATCGTGAACG	1740
GGAGAGAGAG	CGGGAACGGG	ATCCGTCGTC	ACTACTGCCG	CCACCGGCTA	TGGTGCCGCA	1800
GCAAAGACGA	CAGGATGGTG	GACATCAGCG	CTCGTCCTCA	GTGAGCGGAG	TGCCAGGCAG	1860
CAGCTCTTCG	TCGTCTTCCT	CCAGTCACAA	GATGCCAAAT	TACCCTGGTG	GCATGCCGCC	1920
CGAAGCTCAT	CCGGATCACA	AGTCAAAGCA	GCCGGGCTAT	AACAATCGAA	TGCCCTCAAG	1980
TCACCAGCGT	AGTAGTAGCA	GTGGACTCGG	TTCCTCGGGA	AGTGGCAGCC	AGCACAGCAG	2040
CTCATCCTCG	TCGTCTTCAA	GCCAGCAGCC	TGGCCGACCG	TCTATGCCCG	TGGACTATCA	2100
CAAATCCTCT	CGCGGCATGC	CGCCGGTAGG	CGTGGGCATG	CCACCTCACG	GCAGCCACAA	2160
GATGACTTCG	GGCTCCAAGC	CTCAACAGCC	GCAGCAGCAG	CCGGTCCCAC	ATCCATCCGC	2220
CTCTAATTCC	TCTGCATCGG	GCATGTCCTC	CAAGGATAAA	TCCCAGAGCA	ACAAAATGTA	2280
TCCGAACGCA	CCGCCGCCAT	ACAGTAATAG	TGCCCCTCAA	AACCCGCTGA	TGTCGCGTGG	2340
TGGATATCCA	GGCGCTAGCA	ATGGATCCC	GCCCCCGCCT	CCCGCCGGAT	ACGGCGGCCA	2400
TCGCAGCAAA	TCCGGCTCCA	CCGTCCATGG	CATGCCGCAT	TTCGAGCAGC	AATTGCCCTA	2460
TTCCCAGAGC	CAGAGCTACG	GCCACATGCA	GCAGCAGCCA	GTGCCTCAGT	CTCAGCAGCA	2520
ACAGATGCCT	CCGGAGGCAT	CCCAGCACTC	GTTGCAGTCC	AAGAACTCGC	TCTTCAGTCC	2580
AGAGTGGCCA	GACATTAAAA	AGGAGCCAT	GTCGCAGTCG	CAACCACAGC	TTTTTAACGG	2640
TTTGCTACCC	CCTCCTGCGC	CTCCCGGCCA	CGATTACAAG	CTAAATAGCC	ATCCGCGCGA	2700
CAAAGAAAGT	CCCAAGAAAG	AGCGACTAAC	GCCAACCAA	AAGGATAAGC	ACCGTCCTGT	2760
AATGCCCTCA	ATGGGCAGTG	GGAACAGTTC	CTCCGGCTCG	GGATCATCAA	AGCCGATGCT	2820
ACCGCCTCAC	AAGAACGAGA	TACCCCATGG	CGGGGACCTG	TTGACCAATC	CTGGAGAGAG	2880
TGGAAGCCTA	AAACGGCCA	ACGAGATCTC	GGGAAGTCAG	TATGGACTAA	ATAAGCTGGA	2940
TGAAATAGAT	AACAGTAATA	TGCCTCGAGA	AAAGCTTCGC	AAGCTGGACA	CTACAACCTGG	3000
ACTACCAACT	TATCCGAATT	ATGAGGAGAA	ACACACGCCT	CTGAATATGT	CCAACGGAAT	3060
CGAGACAACG	CCGGATCTGG	TGCGCAGTTT	GCTAAAGGAG	AGTCTGTGTC	CATCGAACGC	3120

TTCGCTCCTG AAACCGGATG CCTTGACTAT GCCTGGCCTG AAACCACCGG CCGAACTACT	3180
TGAGCCCATG CCCGCACCAG CGACAATCAA GAAAGAACAG GGAATAACTC CGATGACCAG	3240
TTTGGCTAGT GGGCCCGCAC CCATGGATTG GGAAGTACCC ACTAACACAG CGGGAGAGAT	3300
TAAGGAGGAA AGCAGCAGCA AGTCCGAAAA GAAAAAGAAG AAGGATAAAC ACAAACACAA	3360
GGAGAAGGAC AAGTCCAAGG ACAAGACGGA AAAGGAGGAG CGTAAGAACG ACAAGAGGGA	3420
CAAGCAGAAG GATCGTAGCG GCAGCGGTGG CAGCAAGGAC AGTTCTCTTC CCAATGAGCC	3480
TCTGAAGATG GTTATCAAGA ATCCCCAACGG CAGCCTGCAG GCCGGTGCCT CAGCTCCCAT	3540
TAAACTTAAG ATCAGCAAAA ATAAGGTTGA ACCCAATAAC TACTCTGCAG CGGCGGGTCT	3600
GCCTGGCGCA ATCGGATATG GCTTGCCTCC AACTACGGCT ACCACCACAT CCGCTTCGAT	3660
CGGAGCAGCT GCTCCTGTTG TGCCTCCTTA TGGTGCCGGC GGTGGTGGCT ACAGCTCATC	3720
GGGCGGCAGC AGTTCCGGTG GCAGCAGCAA GAAAAAGCAC AGCGATCGTG ACCGCGACAA	3780
GGAGAGCAAA AAGAATAAGA GCCAAGACTA CGCGAAGTAC AATGGCGCTG GTGGCGGCAT	3840
CTTTAATCCC CTTGGCGGTG CTGGCGCCGC ACCCAATATG TCTGGAGGAA TGGGCGCCCC	3900
CATGTCTACT GCTGTACCAAC CATCCATGCT GTTGGCGCCC ACCGGTGCAG TACCACCCTC	3960
TGCCGCTGGG CTGGCACCGC CTCCCATGCC CGTCTACAAC AAGAAGTAGT GGTAGCGGTC	4020
AGAGGGTTAT TCTTAAGTCG TACGTTTGA TATATGTATA GAACCTCAGT AAGTCCGATT	4080
GTAGTATAGT TGTTAGGATT GTTAGTGAGA TGCATTATTG ATTTTAGTTA AGCACATAGA	4140
TAAAACCTCCA AATTGGAAGT GAAACCGGAT GCGCAGATCG AAGAAGAATG GAAGTAGATG	4200
TCGCGATGGG GCTGGACGTA AAAGCAGTAC TCAAATCGCG AAAACTTTG TACAGCATT	4260
ATTAGTTTAT AACTATAATA AATAGCATAAC ATATAAGCCC AAAAAAAAAA AAAAAAAAAA	4320
AAAAAAAAA	4328

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1097 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Ser Leu Leu Ala Thr Pro Met Pro Gln Ala Ala Thr Ala Ser Ser
1 5 10 15

Ser Ser Ser Ala Ser Ala Ala Ser Ala Ser Gly Ile Pro Ile Thr
20 25 30

Ala Asn Asn Asn Leu Pro Phe Glu Lys Asp Lys Ile Trp Tyr Phe Ser
35 40 45

Asn Asp Gln Leu Ala Asn Leu Pro Ser Arg Arg Cys Gly Ile Lys Gly
50 55 60

Asp Asp Glu Leu Gln Tyr Arg Gln Met Thr Ala Tyr Leu Ile Gln Glu
65 70 75 80

Met Gly Gln Arg Leu Gln Val Ser Gln Leu Cys Ile Asn Thr Ala Ile
85 90 95

Val Tyr Met His Arg Phe Tyr Ala Phe His Ser Phe Thr His Phe His
100 105 110

Arg Asn Ser Met Ala Ser Ala Ser Leu Phe Leu Ala Ala Lys Val Glu
115 120 125

Glu Gln Pro Arg Lys Leu Glu His Val Ile Arg Ala Ala Asn Lys Cys
130 135 140

Leu Pro Pro Thr Thr Glu Gln Asn Tyr Ala Glu Leu Ala Gln Glu Leu
145 150 155 160

Val Phe Asn Glu Asn Val Leu Leu Gln Thr Leu Gly Phe Asp Val Ala
165 170 175

Ile Asp His Pro His Thr His Val Val Arg Thr Cys Gln Leu Val Lys
180 185 190

Ala Cys Lys Asp Leu Ala Gln Thr Ser Tyr Phe Leu Ala Ser Asn Ser
195 200 205

Leu His Leu Thr Ser Met Cys Leu Gln Tyr Arg Pro Thr Val Val Ala
210 215 220

Cys Phe Cys Ile Tyr Leu Ala Cys Lys Trp Ser Arg Trp Glu Ile Pro
225 230 235 240

Gln Ser Thr Glu Gly Lys His Trp Phe Tyr Tyr Val Asp Lys Thr Val
245 250 255

Ser Leu Asp Leu Leu Lys Gln Leu Thr Asp Glu Phe Ile Ala Ile Tyr
260 265 270

DRAFT

Glu Lys Ser Pro Ala Arg Leu Lys Ser Lys Leu Asn Ser Ile Lys Ala
275 280 285

Ile Ala Gln Gly Ala Ser Asn Arg Thr Ala Asn Ser Lys Asp Lys Pro
290 295 300

Lys Glu Asp Trp Lys Ile Thr Glu Met Met Lys Gly Tyr His Ser Asn
305 310 315 320

Ile Thr Thr Pro Pro Glu Leu Leu Asn Gly Asn Asp Ser Arg Asp Arg
325 330 335

Asp Arg Asp Arg Glu Arg Glu Arg Glu Arg Asp Pro Ser Ser
340 345 350

Leu Leu Pro Pro Pro Ala Met Val Pro Gln Gln Arg Arg Gln Asp Gly
355 360 365

Gly His Gln Arg Ser Ser Ser Val Ser Gly Val Pro Gly Ser Ser Ser
370 375 380

Ser Ser Ser Ser Ser His Lys Met Pro Asn Tyr Pro Gly Gly Met
385 390 395 400

Pro Pro Glu Ala His Pro Asp His Lys Ser Lys Gln Pro Gly Tyr Asn
405 410 415

Asn Arg Met Pro Ser Ser His Gln Arg Ser Ser Ser Ser Gly Leu Gly
420 425 430

Ser Ser Gly Ser Gly Ser Gln His Ser Ser Ser Ser Ser Ser Ser Ser
435 440 445

Ser Gln Gln Pro Gly Arg Pro Ser Met Pro Val Asp Tyr His Lys Ser
450 455 460

Ser Arg Gly Met Pro Pro Val Gly Val Gly Met Pro Pro His Gly Ser
465 470 475 480

His Lys Met Thr Ser Gly Ser Lys Pro Gln Gln Pro Gln Gln Pro
485 490 495

Val Pro His Pro Ser Ala Ser Asn Ser Ser Ala Ser Gly Met Ser Ser
500 505 510

Lys Asp Lys Ser Gln Ser Asn Lys Met Tyr Pro Asn Ala Pro Pro Pro
515 520 525

Tyr Ser Asn Ser Ala Pro Gln Asn Pro Leu Met Ser Arg Gly Gly Tyr
530 535 540

Pro Gly Ala Ser Asn Gly Ser Gln Pro Pro Pro Pro Ala Gly Tyr Gly
545 550 555 560

Gly His Arg Ser Lys Ser Gly Ser Thr Val His Gly Met Pro His Phe
565 570 575

Glu Gln Gln Leu Pro Tyr Ser Gln Ser Gln Ser Tyr Gly His Met Gln
580 585 590

Gln Gln Pro Val Pro Gln Ser Gln Gln Gln Met Pro Pro Glu Ala
595 600 605

Ser Gln His Ser Leu Gln Ser Lys Asn Ser Leu Phe Ser Pro Glu Trp
610 615 620

Pro Asp Ile Lys Lys Glu Pro Met Ser Gln Ser Gln Pro Gln Leu Phe
625 630 635 640

Asn Gly Leu Leu Pro Pro Ala Pro Pro Gly His Asp Tyr Lys Leu
645 650 655

Asn Ser His Pro Arg Asp Lys Glu Ser Pro Lys Lys Glu Arg Leu Thr
660 665 670

Pro Thr Lys Lys Asp Lys His Arg Pro Val Met Pro Pro Met Gly Ser
675 680 685

Gly Asn Ser Ser Ser Gly Ser Gly Ser Ser Lys Pro Met Leu Pro Pro
690 695 700

His Lys Lys Gln Ile Pro His Gly Gly Asp Leu Leu Thr Asn Pro Gly
705 710 715 720

Glu Ser Gly Ser Leu Lys Arg Pro Asn Glu Ile Ser Gly Ser Gln Tyr
725 730 735

Gly Leu Asn Lys Leu Asp Glu Ile Asp Asn Ser Asn Met Pro Arg Glu
740 745 750

Lys Leu Arg Lys Leu Asp Thr Thr Gly Leu Pro Thr Tyr Pro Asn
755 760 765

Tyr Glu Glu Lys His Thr Pro Leu Asn Met Ser Asn Gly Ile Glu Thr
770 775 780

Thr Pro Asp Leu Val Arg Ser Leu Leu Lys Glu Ser Leu Cys Pro Ser
785 790 795 800

Asn Ala Ser Leu Leu Lys Pro Asp Ala Leu Thr Met Pro Gly Leu Lys
805 810 815

Pro Pro Ala Glu Leu Leu Glu Pro Met Pro Ala Pro Ala Thr Ile Lys
820 825 830

Lys Glu Gln Gly Ile Thr Pro Met Thr Ser Leu Ala Ser Gly Pro Ala
835 840 845

Pro Met Asp Leu Glu Val Pro Thr Lys Gln Ala Gly Glu Ile Lys Glu
 850 855 860

 Glu Ser Ser Ser Lys Ser Glu Lys Lys Lys Lys Lys Asp Lys His Lys
 865 870 875 880

 His Lys Glu Lys Asp Lys Ser Lys Asp Lys Thr Glu Lys Glu Glu Arg
 885 890 895

 Lys Lys His Lys Arg Asp Lys Gln Lys Asp Arg Ser Gly Ser Gly Gly
 900 905 910

 Ser Lys Asp Ser Ser Leu Pro Asn Glu Pro Leu Lys Met Val Ile Lys
 915 920 925

 Asn Pro Asn Gly Ser Leu Gln Ala Gly Ala Ser Ala Pro Ile Lys Leu
 930 935 940

 Lys Ile Ser Lys Asn Lys Val Glu Pro Asn Asn Tyr Ser Ala Ala Ala
 945 950 955 960

 Gly Leu Pro Gly Ala Ile Gly Tyr Gly Leu Pro Pro Thr Thr Ala Thr
 965 970 975

 Thr Thr Ser Ala Ser Ile Gly Ala Ala Ala Pro Val Leu Pro Pro Tyr
 980 985 990

 Gly Ala Gly Gly Gly Tyr Ser Ser Ser Gly Gly Ser Ser Ser Gly
 995 1000 1005

 Gly Ser Ser Lys Lys His Ser Asp Arg Asp Arg Asp Lys Glu Ser
 1010 1015 1020

 Lys Lys Asn Lys Ser Gln Asp Tyr Ala Lys Tyr Asn Gly Ala Gly Gly
 1025 1030 1035 1040

 Gly Ile Phe Asn Pro Leu Gly Gly Ala Gly Ala Ala Pro Asn Met Ser
 1045 1050 1055

 Gly Gly Met Gly Ala Pro Met Ser Thr Ala Val Pro Pro Ser Met Leu
 1060 1065 1070

 Leu Ala Pro Thr Gly Ala Val Pro Pro Ser Ala Ala Gly Leu Ala Pro
 1075 1080 1085

 Pro Pro Met Pro Val Tyr Asn Lys Lys
 1090 1095

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1119 base pairs
 - (B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 1..1116

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATG GCA AAG CAG TAC GAC TCG GTG GAG TGC CCT TTT TGT GAT GAA GTT	48
Met Ala Lys Gln Tyr Asp Ser Val Glu Cys Pro Phe Cys Asp Glu Val	
1 5 10 15	
TCC AAA TAC GAG AAG CTC GCC AAG ATC GGC CAA GGC ACC TTC GGG GAG	96
Ser Lys Tyr Glu Lys Leu Ala Lys Ile Gly Gln Gly Thr Phe Gly Glu	
20 25 30	
GTG TTC AAG GCC AGG CAC CGC AAG ACC GGC CAG AAG GTG GCT CTG AAG	144
Val Phe Lys Ala Arg His Arg Lys Thr Gly Gln Lys Val Ala Leu Lys	
35 40 45	
AAG GTG CTG ATG GAA AAC GAG AAG GAG GGG TTC CCC ATT ACA GCC TTG	192
Lys Val Leu Met Glu Asn Glu Lys Glu Gly Phe Pro Ile Thr Ala Leu	
50 55 60	
CGG GAG ATC AAG ATC CTT CAG CTT CTA AAA CAC GAG AAT GTG GTC AAC	240
Arg Glu Ile Lys Ile Leu Gln Leu Leu Lys His Glu Asn Val Val Asn	
65 70 75 80	
TTG ATT GAG ATT TGT CGA ACC AAA GCT TCC CCC TAT AAC CGC TGC AAG	288
Leu Ile Glu Ile Cys Arg Thr Lys Ala Ser Pro Tyr Asn Arg Cys Lys	
85 90 95	
GGT AGT ATA TAC CTG GTG TTC GAC TTC TGC GAG CAT GAC CTT GCT GGG	336
Gly Ser Ile Tyr Leu Val Phe Asp Phe Cys Glu His Asp Leu Ala Gly	
100 105 110	
CTG TTG AGC AAT GTT TTG GTC AAG TTC ACG CTG TCT GAG ATC AAG AGG	384
Leu Leu Ser Asn Val Leu Val Lys Phe Thr Leu Ser Glu Ile Lys Arg	
115 120 125	
GTG ATG CAG ATG CTG CTT AAC GGC CTC TAC TAC ATC CAC AGA AAC AAG	432
Val Met Gln Met Leu Leu Asn Gly Leu Tyr Tyr Ile His Arg Asn Lys	
130 135 140	
ATC CTG CAT AGG GAC ATG AAG GCT GCT AAT GTG CTT ATC ACT CGT GAT	480
Ile Leu His Arg Asp Met Lys Ala Ala Asn Val Leu Ile Thr Arg Asp	
145 150 155 160	
GGG GTC CTG AAG CTG GCA GAC TTT GGG CTG GCC CGG GCC TTC AGC CTG	528
Gly Val Leu Lys Leu Ala Asp Phe Gly Leu Ala Arg Ala Phe Ser Leu	
165 170 175	

DRAFT - 10/2009

GCC AAG AAC AGC CAG CCC AAC CGC TAC ACC AAC CGT GTG GTG ACA CTC Ala Lys Asn Ser Gln Pro Asn Arg Tyr Thr Asn Arg Val Val Thr Leu 180 185 190	576
TGG TAC CGG CCC CCG GAG CTG TTG CTC GGG GAG CGG GAC TAC GGC CCC Trp Tyr Arg Pro Pro Glu Leu Leu Gly Glu Arg Asp Tyr Gly Pro 195 200 205	624
CCC ATT GAC CTG TGG GGT GCT GGG TGC ATC ATG GCA GAG ATG TGG ACC Pro Ile Asp Leu Trp Gly Ala Gly Cys Ile Met Ala Glu Met Trp Thr 210 215 220	672
CGC AGC CCC ATC ATG CAG GGC AAC ACG GAG CAG CAC CAA CTC GCC CTC Arg Ser Pro Ile Met Gln Gly Asn Thr Glu Gln His Gln Leu Ala Leu 225 230 235 240	720
ATC AGT CAG CTC TGC GGC TCC ATC ACC CCT GAG GTG TGG CCA AAC GTG Ile Ser Gln Leu Cys Gly Ser Ile Thr Pro Glu Val Trp Pro Asn Val 245 250 255	768
GAC AAC TAT GAG CTG TAC GAA AAG CTG GAG CTG GTC AAG GGC CAG AAG Asp Asn Tyr Glu Leu Tyr Glu Lys Leu Glu Leu Val Lys Gly Gln Lys 260 265 270	816
CGG AAG GTG AAG GAC AGG CTG AAG GCC TAT GTG CGT GAC CCA TAC GCA Arg Lys Val Lys Asp Arg Leu Lys Ala Tyr Val Arg Asp Pro Tyr Ala 275 280 285	864
CTG GAC CTC ATC GAC AAG CTG CTG GTG CTG GAC CCT GCC CAG CGC ATC Leu Asp Leu Ile Asp Lys Leu Leu Val Leu Asp Pro Ala Gln Arg Ile 290 295 300	912
GAC AGC GAT GAC GCC CTC AAC CAC GAC TTC TTC TGG TCC GAC CCC ATG Asp Ser Asp Asp Ala Leu Asn His Asp Phe Phe Trp Ser Asp Pro Met 305 310 315 320	960
CCC TCC GAC CTC AAG GGC ATG CTC TCC ACC CAC CTG ACG TCC ATG TTC Pro Ser Asp Leu Lys Gly Met Leu Ser Thr His Leu Thr Ser Met Phe 325 330 335	1008
GAG TAC TTG GCA CCA CCG CGC CGG AAG GGC AGC CAG ATC ACC CAG CAG Glu Tyr Leu Ala Pro Pro Arg Arg Lys Gly Ser Gln Ile Thr Gln Gln 340 345 350	1056
TCC ACC AAC CAG AGT CGC AAT CCC GCC ACC ACC AAC CAG ACG GAG TTT Ser Thr Asn Gln Ser Arg Asn Pro Ala Thr Thr Asn Gln Thr Glu Phe 355 360 365	1104
GAG CGC GTC TTC TGA Glu Arg Val Phe 370	1119

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 372 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Ala Lys Gln Tyr Asp Ser Val Glu Cys Pro Phe Cys Asp Glu Val
1 5 10 15

Ser Lys Tyr Glu Lys Leu Ala Lys Ile Gly Gln Gly Thr Phe Gly Glu
20 25 30

Val Phe Lys Ala Arg His Arg Lys Thr Gly Gln Lys Val Ala Leu Lys
35 40 45

Lys Val Leu Met Glu Asn Glu Lys Glu Gly Phe Pro Ile Thr Ala Leu
50 55 60

Arg Glu Ile Lys Ile Leu Gln Leu Leu Lys His Glu Asn Val Val Asn
65 70 75 80

Leu Ile Glu Ile Cys Arg Thr Lys Ala Ser Pro Tyr Asn Arg Cys Lys
85 90 95

Gly Ser Ile Tyr Leu Val Phe Asp Phe Cys Glu His Asp Leu Ala Gly
100 105 110

Leu Leu Ser Asn Val Leu Val Lys Phe Thr Leu Ser Glu Ile Lys Arg
115 120 125

Val Met Gln Met Leu Leu Asn Gly Leu Tyr Tyr Ile His Arg Asn Lys
130 135 140

Ile Leu His Arg Asp Met Lys Ala Ala Asn Val Leu Ile Thr Arg Asp
145 150 155 160

Gly Val Leu Lys Leu Ala Asp Phe Gly Leu Ala Arg Ala Phe Ser Leu
165 170 175

Ala Lys Asn Ser Gln Pro Asn Arg Tyr Thr Asn Arg Val Val Thr Leu
180 185 190

Trp Tyr Arg Pro Pro Glu Leu Leu Gly Glu Arg Asp Tyr Gly Pro
195 200 205

Pro Ile Asp Leu Trp Gly Ala Gly Cys Ile Met Ala Glu Met Trp Thr
210 215 220

DRAFT - DO NOT DISTRIBUTE

Arg Ser Pro Ile Met Gln Gly Asn Thr Glu Gln His Gln Leu Ala Leu
225 230 235 240

Ile Ser Gln Leu Cys Gly Ser Ile Thr Pro Glu Val Trp Pro Asn Val
245 250 255

Asp Asn Tyr Glu Leu Tyr Glu Lys Leu Glu Leu Val Lys Gly Gln Lys
260 265 270

Arg Lys Val Lys Asp Arg Leu Lys Ala Tyr Val Arg Asp Pro Tyr Ala
275 280 285

Leu Asp Leu Ile Asp Lys Leu Leu Val Leu Asp Pro Ala Gln Arg Ile
290 295 300

Asp Ser Asp Asp Ala Leu Asn His Asp Phe Phe Trp Ser Asp Pro Met
305 310 315 320

Pro Ser Asp Leu Lys Gly Met Leu Ser Thr His Leu Thr Ser Met Phe
325 330 335

Glu Tyr Leu Ala Pro Pro Arg Arg Lys Gly Ser Gln Ile Thr Gln Gln
340 345 350

Ser Thr Asn Gln Ser Arg Asn Pro Ala Thr Thr Asn Gln Thr Glu Phe
355 360 365

Glu Arg Val Phe
370

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ACGAATTCCA CACAATCCAA AGATC

25

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CAGAATTCCCT ATTGCCGATC CCCAGA

26

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: one-of(8, 14)
- (D) OTHER INFORMATION: /mod_base= OTHER

/note= "N = A or C or G or T"

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 12
- (D) OTHER INFORMATION: /mod_base= OTHER

/note= "Y = C or T"

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: one-of(17, 20)
- (D) OTHER INFORMATION: /mod_base= OTHER

/note= "R = A or G"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGAATTCNAT GYTNCARCAR CC

22

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: one-of(13, 16, 19, 22, 25)
- (D) OTHER INFORMATION: /mod_base= OTHER

/note= "R = A or G"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AACTGCAGTC CARAARAART CRTGRTT

27

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TGTCAAGGAT CAAACCGGCT GTGAT

25

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CGAATTCCAA GAAACGCATC GATGC

25

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AGACCTGCCA AATCGTGT

18

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AGAAGGTGGA TCTGTAACCA TTCGT

25

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GGAATTCAGA TCTCGATCAG ATTCA

25

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TTACTACTCG AGCTACCAAA CCCGGTC

27

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 28 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TAAGCAAGCT TCTATGGCGC ACATGTCC

28

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TTACTACTCG AGCTACCAAA CCCGGTC

27

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: one-of(13, 16, 22)
- (D) OTHER INFORMATION: /mod_base= OTHER

/note= "Y = C or T"

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 17
- (D) OTHER INFORMATION: /mod_base= OTHER

/note= "W = A or T"

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 18
- (D) OTHER INFORMATION: /mod_base= OTHER

/note= "S = C or G"

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 19
- (D) OTHER INFORMATION: /mod_base= OTHER

/note= "N = A or C or G or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GGAATTCTGG TAYTTYWSNA AYGA

24

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 11
- (D) OTHER INFORMATION: /mod_base= OTHER

/note= "Y = C or T"

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 14

(D) OTHER INFORMATION: /mod_base= OTHER
/note= "R = A or G"

(ix) FEATURE:

(A) NAME/KEY: modified_base
(B) LOCATION: one-of(17, 20)
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "N = A or C or G or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CGGGATCCTG YTCRAANGGN GGCAT

25

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: modified_base
(B) LOCATION: one-of(11, 14, 20)
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "N = A or C or G or T"

(ix) FEATURE:

(A) NAME/KEY: modified_base
(B) LOCATION: 23
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "R = A or G"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CGGGATCCAA NGGNNGCATN CCRT

24

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

ATCACGACAC CACCAGAGCT GTTA

24

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CGAATTCAAGA TCGTGAAACGG GA

22

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CGAATTCAAGG CGCTAGCAAT G

21

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GAAAGGCGTA GAACCGA

17

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GCTGACCCAT TTCCTGTATC AGATAG

26

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GGAATTCTTC TGCTTGGCGA AT

22

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GGGAATTCGA GGTTCTATAAC ATAT

24

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

CTGTGTGAAT GGAATCTGTG ATGTG

25

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 28 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

TATCCCGGGT CATATGAGTC TCCTAGCC

28

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 13 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Met Leu Gln Gln Pro Ser Gly Ser Thr Pro Ser Asn Val
1 5 10

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 17 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Ala Asp Thr Ala Leu Asn His Asp Phe Phe Trp Thr Asp Pro Met Pro
1 5 10 15
Ser

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Met Leu Gln Gln Pro
1 5

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Asn His Asp Phe Phe Trp Thr
1 5

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Ser Pro Glu Trp Pro Asp Ile
1 5

(2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Trp Tyr Phe Ser Asn Asp Gln Leu Ala Asn Ser Pro Ser Arg
1 5 10

(2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Thr Val His Gly Met Pro Pro Phe Glu Gln Gln Leu Pro Tyr
1 5 10

(2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Trp Tyr Phe Ser Asn Asp
1 5

(2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Met Pro Pro Phe Glu Gln
1 5

(2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

His Gly Met Pro Pro Phe
1 5

(2) INFORMATION FOR SEQ ID NO:41:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GCAGGATCCA GAATTCCATA TGGCAAAGCA GTACGACTCG G

41

(2) INFORMATION FOR SEQ ID NO:42:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

CAGTACTCGA GTTATCAGAA GACGCGCTCA AAC

33

(2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4528 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GGGGGGGGGGGG	GGGTGAATGA	AGGAGCGGGC	GGAGGAGGAA	TTGTCATGGC	GTCGGGCCGT	60
GGAGCTTCTT	CTCGCTGGTT	CTTTACTCGG	GAACAGCTGG	AGAACACGCC	GAGCCGCCGC	120
TGCGGAGTGG	AGGC GGATAA	AGAGCTCTCG	TGCCGCCAGC	AGGCGGCCAA	CCTCATCCAG	180
GAGATGGGAC	AGCGTCTCAA	TGTCTCTCAG	CTTACAATAA	ACACTGCGAT	TGTTTATATG	240
CACAGGTTTT	ATATGCACCA	TTCTTCACC	AAATTCAACA	AAAATATAAT	ATCGTCTACT	300
GCATTATTTT	TGGCTGAAA	AGTGGAAAGAA	CAGGCTCGAA	AACTTGAACA	TGTTATCAA	360
GTAGCACATG	CTTGTCTTCA	TCCTCTAGAG	CCACTGCTGG	ATACTAAATG	TGATGCTTAC	420
CTTCAACAGA	CTCAAGAACT	GGTTATACTT	GAAACCATAA	TGCTACAAAC	TCTAGGTTTT	480
GAGATCACCA	TTGAACACCC	ACACACAGAT	GTGGTGAAAT	GTACCCAGTT	AGTAAGAGCA	540
AGCAAGGATT	TGGCACAGAC	ATCCTATTTC	ATGGCTACCA	ACAGTCTGCA	TCTTACAACC	600
TTCTGTCTTC	AGTACAAACC	AACAGTGATA	GCATGTGTAT	GCATTCA	GGCTTGCAA	660
TGGTCCAATT	GGGAGATCCC	TGTATCAACT	GATGGAAAGC	ATTGGTGGGA	ATATGTGGAT	720
CCTACAGTTA	CTCTAGAATT	ATTAGATGAG	CTAACACATG	AGTTTCTACA	AATATTGGAG	780
AAAACGCCTA	ATAGGTTGAA	GAAGATTCGA	AACTGGAGGG	CTAACAGGC	AGCTAGGAAA	840
CCAAAAGTAG	ATGGACAGGT	ATCAGAGACA	CCACTCTTG	GTTCATCTT	GGTCCAGAAT	900
TCCATTTAG	TAGATAGTGT	CACTGGTGTG	CCTACAAACC	CAAGTTTCA	GAAACCATCT	960
ACATCAGCAT	TCCCTGCGCC	AGTACCTCTA	AATTCAAGGAA	ATATTTCTGT	TCAAGACAGC	1020
CATACATCTG	ATAATTGTC	AATGCTAGCA	ACAGGAATGC	CAAGTACTTC	ATACGGTTA	1080

2020-12-29

TCATCACACC AGGAATGGCC TCAACATCAA GACTCAGCAA GGACAGAACAA GCTATATTCA	1140
CAGAAACAGG AGACATCTT GTCTGGTAGC CAGTACAACA TCAACTTCCA GCAGGGACCT	1200
TCTATATCAC TGCATTCAGG ATTACATCAC AGACCTGACA AAATTCAGA TCATTCTTCT	1260
GTAAAGCAAG AATATACTCA TAAAGCAGGG AGCAGTAAAC ACCATGGGCC AATTTCCACT	1320
ACTCCAGGAA TAATTCTCA GAAAATGTCT TTAGATAAAAT ATAGAGAAAA GCGTAAACTA	1380
GAAACTCTTG ATCTCGATGT AAGGGATCAT TATATAGCTG CCCAGGTAGA ACAGCAGCAC	1440
AAACAAGGGC AGTCACAGGC AGCCAGCAGC AGTTCTGTTA CTTCTCCCCT TAAAATGAAA	1500
ATACCTATCG CAAATACTGA AAAATACATG GCAGATAAAA AGGAAAAGAG TGGGTCACTG	1560
AAATTACGGA TTCCAATACC ACCCACTGAT AAAAGGCCA GTAAAGAAGA ACTGAAAATG	1620
AAAATAAAAG TTTCTCTTC AGAAAGACAC AGCTCTTCTG ATGAAGGCAG TGGGAAAAGC	1680
AAACATTCAA GCCCACATAT TAGCAGAGAC CATAAGGAGA AGCACAAGGA GCATCCTTCA	1740
AGCCGCCACC ACACCAGCAG CCACAAGCAT TCCCACTCGC ATAGTGGCAG CAGCAGCGGT	1800
GGCAGTAAAC ACAGTGCCGA CGGAATACCA CCCACTGTT TGAGGAGTCC TGTTGGCCTG	1860
AGCAGTGATG GCATTTCTC TAGCTCCAGC TCTTCAAGGA AGAGGCTGCA TGTCAATGAT	1920
GCATCTCACA ACCACCACTC CAAAATGAGC AAAAGTTCCA AAAGTTCAGG TGGGCTACGG	1980
ACATCTCAGC ACCTCGTGAA ACTGGACAAG AAGCCAGTGG AGACCAACGG TCCTGATGCC	2040
AATCACGAGT ACAGTACAAG CAGCCAGCAT ATGGACTACA AAGACACATT CGACATGCTG	2100
GACTCACTGT TAAGTGCCCA AGGAATGAAC ATGTAATAAT TTGTTTAGGT CAATTTTCC	2160
TTTACTTTTT TAATTAAAA ATTGTTAGAA TGGAAAAATT CCTTCTGATC TAGCAGTGGT	2220
AACCCCTGCT GTTGCTGCCA CTGCTTCAAT ATTTGTAAGT GCTACTTTAT TCTTCATTCT	2280
GAAAAGAAGA GATTATAGTA AACAAAGTCTT TATCTCCACA TATGATAGTG TTATAAATAC	2340
TGTAAAGGCA TGGAAGGTGC AAAACTCAGT ATTTCTACAA TTGCAGCTAA GAACATTAGG	2400
ATGAATGGCT GGCTGCTTCT AGGAATATAA GATGCCTCAA GCATTCATTA TTTATGATT	2460
GAATACTGTA GCTATTTTT GTTGCTTGGC TTTTGAATGA GTGTAAATTG TTTTCTTTG	2520
TGTATTTATA CTTGTATGTA TGATTGAT GTTCAATGA TAAAGGGATA AAACAGTATA	2580
CTGACAACTG TTTACAAGAA AGTGGAGAAA ATGTAACACA TTTTGTATGT TTAGATATTA	2640
CCGTAAATAC TCAGGATTGG AGCTGCTTGT AAGTATAACA ATATACAGAA TACTTTATTT	2700

TATCTTGTCA GAGTTCCATC ACTATCTAAA ACAAAAGGTGC AATTTTTAT GTTAACCTTA	2760
AATCTAGCCC TTACTGGAAG CCACTGATAG GGACATTAC TACCAGATGT GTGCAGTGCA	2820
GCAGATGGTC ATATAACACT GTGAGGCAC TGAATTTGCC TTCAGAGGTT CTGACCAGAT	2880
TGGCTGCTGA AATAGCCCCT AACTTTCTGA AGGCTTGAAG AGGAAAAAAT AAAGTTTACA	2940
TACTCTTGAT GGAAGTGCAT TTAAATGTTT GTTGGCTTGT TGCAGTTCTA TGAAACAGAG	3000
CTGTTAATAA TGGTTATGTG GATTACTGTG ATTTGAAAAC TAAATTACACA ATAACCTTACC	3060
TAGTAGAGAT TTAGTGAGTT GTTCCCTTA AAGAATTAA CACTACATAT TTTAATAGTA	3120
AACAGGGTCA CTTTCCTTA GCATTCAAGAA TGACACCATA TTCTTAAATA TACTCCTTCC	3180
CTGAAGCGTG TTTGTGTGTG ATGCCATATT TCTTTTCAG GTAAATGTAG TCTTCCTTAT	3240
AAAAATGAAA TTAAACCTAT GCTCTCAATT CTTTTATATT CTAACAATAA ATAAAAAAAGA	3300
AAAGATTACT GACTGTGCAT TGTACCTGTA TTTATAGTTT ATGGTTATCA GAAGCTCTGT	3360
AAGAAAGAAA AGGTCAGCTC CCAGGCAAAC CAGTAGTGGA GGTTTTACAT TTGTTTGCAC	3420
ATCTCAGTAT ATTTCTGTTG AGGTAAAGTT TGCACAGTCA TCTGACTTCT GATCAAGCAT	3480
TAGATTTAA CTTGTTAGA TTTGCTTA AACACCAGTA ATATGGCTCT TGTTTATCAG	3540
CTAATCTTGA ATTTATTCTG TGGTAAATCT TTTGAGTTGC TGAGTATATT TGAGATTGAT	3600
TGGATTCAAC CTCTTGTGA ACTGAAAACT TAATTTTTC TCTGTATTT TGTTACAAAG	3660
CCACTGATAC GTGCACAATT GTAATTAAGT ATGTTGCAGT TGTAAATATT AGAGTTAAC	3720
CTCATGCTCT ACCTTTATTT AGCAATTACC TAATTTGCCA GTAGCTTAT AATTTTAAA	3780
GATAATTGTT CATTATTTTG TCAATGTTAT TTGAACCTGG GGTACTTAGG AGCCTCTTG	3840
TAGGGACTGT GCCTAGGTAG CATGTCCTAA CATTGTTCT GGTCTTGCAT AACTCAGTA	3900
TCTTGTCAT TATATGTAAC TTTGTTGCTC TGTATGGCAT AATATTGTAT CCATAAACAT	3960
GGTAATTTTG ATACAGTTAT ACTTTACAG TGGTACATAA TCCAAGGACT AGTATAGAAT	4020
TAAGCTGAGT GCAAGATGAG GGAGGGAAGG GCTTTCTTGG TAATTTAGAT GTGAAACCTC	4080
TACAGAGCTA TCATGTAAAA ACTACATGAG GTGGTTGTGC TACTGTATAA TTGGGGGTGA	4140
TAATACCAGG AATTTAATA AGATTTGTA AAGAATATCC AGAAAAGTAG TGAACTTATT	4200
TTCAGTAGGC ATAGAAAACA ATGTGAATAT TTAAGGTCTG TGACTATAGT TAAACTTCAC	4260
TAAGAATTG CAGAATTGTT TTGAGATGTG TGAATAAAGG TAATTTATT GAATCTTCAT	4320

TGGTGCTAAT GTTGGACAGT TAAAAAGATA GCTAGTGTAT ATTGTTATGG GTCAGTACTT	4380
ATTAGTACTT CCAAAATTGA ATTTGAAATG CTATGTATTC ACTTTCACT CTGTAAATGT	4440
AATTCTTTAC AATGACTTTA TTTATTAAAG GGCAGCCAGT TGTCATTTGT AAAAAAAAAA	4500
AAAAAAAAAA AAAGCGGCCG CTGAATTC	4528

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2091 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

ATGGCGTCGG GCCGTGGAGC TTCTTCTCGC TGGTTCTTA CTCGGAAACA GCTGGAGAAC	60
ACGCCGAGCC GCCGCTGCGG AGTGGAGGCG GATAAAGAGC TCTCGTGCG CCAGCAGGCG	120
GCCAACCTCA TCCAGGAGAT GGGACAGCGT CTCAATGTCT CTCAGCTTAC AATAAACACT	180
GCGATTGTTT ATATGCACAG GTTTTATATG CACCATTCTT TCACCAAATT CAACAAAAAT	240
ATAATATCGT CTACTGCATT ATTTTGCT GCAAAAGTGG AAGAACAGGC TCGAAAACCTT	300
GAACATGTTA TCAAAGTAGC ACATGCTTGT CTTCATCCTC TAGAGCCACT GCTGGATACT	360
AAATGTGATG CTTACCTTCA ACAGACTCAA GAACGGTTA TACTTGAAAC CATAATGCTA	420
CAAACCTCTAG GTTTGAGAT CACCATTGAA CACCCACACA CAGATGTGGT GAAATGTACC	480
CAGTTAGTAA GAGCAAGCAA GGATTTGGCA CAGACATCCT ATTTCATGGC TACCAACAGT	540
CTGCATCTTA CAACCTTCTG TCTTCAGTAC AAACCAACAG TGATAGCATG TGTATGCATT	600
CATTTGGCTT GCAAATGGTC CAATTGGAG ATCCCTGTAT CAACTGATGG AAAGCATTGG	660
TGGGAATATG TGGATCCTAC AGTTACTCTA GAATTATTAG ATGAGCTAAC ACATGAGTTT	720
CTACAAATAT TGGAGAAAAC GCCTAATAGG TTGAAGAAGA TTCGAAACTG GAGGGCTAAT	780
CAGGCAGCTA GGAAACCAAA AGTAGATGGA CAGGTATCAG AGACACCACT TCTTGGTTCA	840
TCTTTGGTCC AGAATTCCAT TTTAGTAGAT AGTGTCACTG GTGTGCCTAC AAACCCAAGT	900
TTTCAGAAAC CATCTACATC AGCATTCCCT GCGCCAGTAC CTCTAAATTC AGGAAATATT	960
TCTGTTCAAG ACAGCCATAC ATCTGATAAT TTGTCAATGC TAGCAACAGG AATGCCAAGT	1020
ACTTCATACG GTTTATCATC ACACCAGGAA TGGCCTCAAC ATCAAGACTC AGCAAGGACA	1080

GAACAGCTAT ATTACAGAA ACAGGAGACA TCTTGTCTG GTAGCCAGTA CAACATCAAC 1140
 TTCCAGCAGG GACCTTCTAT ATCACTGCAT TCAGGATTAC ATCACAGACC TGACAAAATT 1200
 TCAGATCATT CTTCTGTTAA GCAAGAATAT ACTCATAAAG CAGGGAGCAG TAAACACCAT 1260
 GGGCCAATT CCACTAATCC AGGAATAATT CCTCAGAAAA TGTCTTAGA TAAATATAGA 1320
 GAAAAGCGTA AACTAGAAC TCTTGATCTC GATGTAAGGG ATCATTATAT AGCTGCCAG 1380
 GTAGAACAGC AGCACAAACA AGGGCAGTCA CAGGCAGCCA GCAGCAGTTC TGTTACTTCT 1440
 CCCATTAAAA TGAAAATACC TATCGCAAAT ACTGAAAAAT ACATGGCAGA TAAAAAGGAA 1500
 AAGAGTGGGT CACTGAAATT ACGGATTCCA ATACCACCCA CTGATAAAAG CGCCAGTAA 1560
 GAAGAACTGA AAATGAAAAT AAAAGTTCT TCTTCAGAAA GACACAGCTC TTCTGATGAA 1620
 GGCAGTGGGA AAAGCAAACA TTCAAGCCC CATATTAGCA GAGACCATAA GGAGAACGAC 1680
 AAGGAGCATIC CTTCAAGCCG CCACCACACC AGCAGCCACA AGCATTCCC CTCGCATAGT 1740
 GGCAGCAGCA GCGGTGGCAG TAAACACAGT GCCGACGGAA TACCACCCAC TGTTCTGAGG 1800
 AGTCCTGTTG GCCTGAGCAG TGATGGCATT TCCTCTAGCT CCAGCTCTTC AAGGAAGAGG 1860
 CTGCATGTCA ATGATGCATC TCACAACCAC CACTCCAAA TGAGCAAAAG TTCCAAAAGT 1920
 TCAGGTGGGC TACGGACATC TCAGCACCTC GTGAAACTGG ACAAGAAGCC AGTGGAGACC 1980
 AACGGTCCTG ATGCCAATCA CGAGTACAGT ACAAGCAGCC AGCATATGGA CTACAAAGAC 2040
 ACATTCGACA TGCTGGACTC ACTGTTAAGT GCCCAAGGAA TGAACATGTA A 2091

(2) INFORMATION FOR SEQ ID NO:45:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 696 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

Met	Ala	Ser	Gly	Arg	Gly	Ala	Ser	Ser	Arg	Trp	Phe	Phe	Thr	Arg	Glu
1															

Gln	Leu	Glu	Asn	Thr	Pro	Ser	Arg	Arg	Cys	Gly	Val	Glu	Ala	Asp	Lys
20															

Glu	Leu	Ser	Cys	Arg	Gln	Gln	Ala	Ala	Asn	Leu	Ile	Gln	Glu	Met	Gly
35															

Gln Arg Leu Asn Val Ser Gln Leu Thr Ile Asn Thr Ala Ile Val Tyr
50 55 60

Met His Arg Phe Tyr Met His His Ser Phe Thr Lys Phe Asn Lys Asn
65 70 75 80

Ile Ile Ser Ser Thr Ala Leu Phe Leu Ala Ala Lys Val Glu Glu Gln
85 90 95

Ala Arg Lys Leu Glu His Val Ile Lys Val Ala His Ala Cys Leu His
100 105 110

Pro Leu Glu Pro Leu Leu Asp Thr Lys Cys Asp Ala Tyr Leu Gln Gln
115 120 125

Thr Gln Glu Leu Val Ile Leu Glu Thr Ile Met Leu Gln Thr Leu Gly
130 135 140

Phe Glu Ile Thr Ile Glu His Pro His Thr Asp Val Val Lys Cys Thr
145 150 155 160

Gln Leu Val Arg Ala Ser Lys Asp Leu Ala Gln Thr Ser Tyr Phe Met
165 170 175

Ala Thr Asn Ser Leu His Leu Thr Thr Phe Cys Leu Gln Tyr Lys Pro
180 185 190

Thr Val Ile Ala Cys Val Cys Ile His Leu Ala Cys Lys Trp Ser Asn
195 200 205

Trp Glu Ile Pro Val Ser Thr Asp Gly Lys His Trp Trp Glu Tyr Val
210 215 220

Asp Pro Thr Val Thr Leu Glu Leu Leu Asp Glu Leu Thr His Glu Phe
225 230 235 240

Leu Gln Ile Leu Glu Lys Thr Pro Asn Arg Leu Lys Lys Ile Arg Asn
245 250 255

Trp Arg Ala Asn Gln Ala Ala Arg Lys Pro Lys Val Asp Gly Gln Val
260 265 270

Ser Glu Thr Pro Leu Leu Gly Ser Ser Leu Val Gln Asn Ser Ile Leu
275 280 285

Val Asp Ser Val Thr Gly Val Pro Thr Asn Pro Ser Phe Gln Lys Pro
290 295 300

Ser Thr Ser Ala Phe Pro Ala Pro Val Pro Leu Asn Ser Gly Asn Ile
305 310 315 320

Ser Val Gln Asp Ser His Thr Ser Asp Asn Leu Ser Met Leu Ala Thr
325 330 335

Gly Met Pro Ser Thr Ser Tyr Gly Leu Ser Ser His Gln Glu Trp Pro
340 345 350

Gln His Gln Asp Ser Ala Arg Thr Glu Gln Leu Tyr Ser Gln Lys Gln
355 360 365

Glu Thr Ser Leu Ser Gly Ser Gln Tyr Asn Ile Asn Phe Gln Gln Gly
370 375 380

Pro Ser Ile Ser Leu His Ser Gly Leu His His Arg Pro Asp Lys Ile
385 390 395 400

Ser Asp His Ser Ser Val Lys Gln Glu Tyr Thr His Lys Ala Gly Ser
405 410 415

Ser Lys His His Gly Pro Ile Ser Thr Thr Pro Gly Ile Ile Pro Gln
420 425 430

Lys Met Ser Leu Asp Lys Tyr Arg Glu Lys Arg Lys Leu Glu Thr Leu
435 440 445

Asp Leu Asp Val Arg Asp His Tyr Ile Ala Ala Gln Val Glu Gln Gln
450 455 460

His Lys Gln Gly Gln Ser Gln Ala Ala Ser Ser Ser Ser Val Thr Ser
465 470 475 480

Pro Ile Lys Met Lys Ile Pro Ile Ala Asn Thr Glu Lys Tyr Met Ala
485 490 495

Asp Lys Lys Glu Lys Ser Gly Ser Leu Lys Leu Arg Ile Pro Ile Pro
500 505 510

Pro Thr Asp Lys Ser Ala Ser Lys Glu Glu Leu Lys Met Lys Ile Lys
515 520 525

Val Ser Ser Ser Glu Arg His Ser Ser Ser Asp Glu Gly Ser Gly Lys
530 535 540

Ser Lys His Ser Ser Pro His Ile Ser Arg Asp His Lys Glu Lys His
545 550 555 560

Lys Glu His Pro Ser Ser Arg His His Thr Ser Ser His Lys His Ser
565 570 575

His Ser His Ser Gly Ser Ser Ser Gly Gly Ser Lys His Ser Ala Asp
580 585 590

Gly Ile Pro Pro Thr Val Leu Arg Ser Pro Val Gly Leu Ser Ser Asp
595 600 605

Gly Ile Ser Ser Ser Ser Ser Ser Arg Lys Arg Leu His Val Asn
610 615 620

Asp Ala Ser His Asn His His Ser Lys Met Ser Lys Ser Ser Lys Ser
 625 630 635 640
 Ser Gly Gly Leu Arg Thr Ser Gln His Leu Val Lys Leu Asp Lys Lys
 645 650 655
 Pro Val Glu Thr Asn Gly Pro Asp Ala Asn His Glu Tyr Ser Thr Ser
 660 665 670
 Ser Gln His Met Asp Tyr Lys Asp Thr Phe Asp Met Leu Asp Ser Leu
 675 680 685
 Leu Ser Ala Gln Gly Met Asn Met
 690 695

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2190 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

ATGGCGTCGG	GCCGTGGAGC	TTCTTCTCGC	TGGTTCTTTA	CTCGGAAACA	GCTGGAGAAC	60
ACGCCGAGCC	GCCGCTGCGG	AGTGGAGGCG	GATAAAGAGC	TCTCGTGCCG	CCAGCAGGCG	120
GCCAACCTCA	TCCAGGAGAT	GGGACAGCGT	CTCAATGTCT	CTCAGCTTAC	AATAAACACT	180
GCGATTGTTT	ATATGCACAG	GTTTTATATG	CACCATTCTT	TCACCAAATT	CAACAAAAAT	240
ATAATATCGT	CTACTGCATT	ATTTTGCGCT	GCAAAAGTGG	AAGAACAGGC	TCGAAAACTT	300
GAACATGTTA	TCAAAGTAGC	ACATGCTTGT	CTTCATCCTC	TAGAGCCACT	GCTGGATACT	360
AAATGTGATG	CTTACCTTCA	ACAGACTCAA	GAACTGGTTA	TACTTGAAAC	CATAATGCTA	420
CAAACCTCTAG	GTTTGAGAT	CACCATTGAA	CACCCACACA	CAGATGTGGT	GAAATGTACC	480
CAGTTAGTAA	GAGCAAGCAA	GGATTTGGCA	CAGACATCCT	ATTTCATGGC	TACCAACAGT	540
CTGCATCTTA	CAACCTTCTG	TCTTCAGTAC	AAACCAACAG	TGATAGCATG	TGTATGCATT	600
CATTTGGCTT	GCAAATGGTG	CAATTGGGAG	ATCCCTGTAT	CAACTGATGG	AAAGCATTGG	660
TGGGAATATG	TGGATCCTAC	AGTTACTCTA	GAATTATTAG	ATGAGCTAAC	ACATGAGTTT	720
CTACAAATAT	TGGAGAAAAC	GCCTAATAGG	TTGAAGAAGA	TTCGAAACTG	GAGGGCTAAT	780
CAGGCAGCTA	GGAAACCAAA	AGTAGATGGA	CAGGTATCAG	AGACACCACT	TCTTGGTTCA	840

TCTTTGGTCC AGAATTCCAT TTTAGTAGAT AGTGTCACTG GTGTGCCTAC AAACCCAAGT 900
 TTTCAGAAAC CATCTACATC AGCATTCCCT GCGCCAGTAC CTCTAAATT AGGAAATATT 960
 TCTGTTCAAG ACAGCCATAC ATCTGATAAT TTGTCAATGC TAGAACAGG AATGCCAAGT 1020
 ACTTCATACG GTTTATCATC ACACCAGGAA TGGCCTCAAC ATCAAGACTC AGCAAGGACA 1080
 GAACAGCTAT ATTACACAGAA ACAGGAGACA TCTTTGTCTG GTAGCCAGTA CAACATCAAC 1140
 TTCCAGCAGG GACCTTCTAT ATCACTGCAT TCAGGATTAC ATCACAGACC TGACAAAATT 1200
 TCAGATCATT CTTCTGTTAA GCAGGAATAT ACTCATAAAG CAGGGAGCAG TAAACACCAT 1260
 GGGCCAATT CCACTACTCC AGGAATAATT CCTCAGAAAAA TGTCTTTAGA TAAATATAGA 1320
 GAAAAGCGTA AACTAGAAC TCTTGATCTC GATGTAAGGG ATCATTATAT AGCTGCCAG 1380
 GTAGAACAGC AGCACAAACA AGGGCAGTCA CAGGCAGCCA GCAGCAGTTC TGTTACTTCT 1440
 CCCATTAAAA TGAAAATACC TATCGCAAAT ACTGAAAAAT ACATGGCAGA TAAAAAGGAA 1500
 AAGAGTGGGT CACTGAAATT ACGGATTCCA ATACCACCCA CTGATAAAAG CGCCAGTAAA 1560
 GAAGAACTGA AAATGAAAAT AAAAGTTCT TCTTCAGAAA GACACAGCTC TTCTGATGAA 1620
 GGCAGTGGGA AAAGCAAACA TTCAAGCCC CATATTAGCA GAGACCATAA GGAGAACAC 1680
 AAGGAGCATE CTTCAAGCCG CCACCACACC AGCAGCCACA AGCATTCCCA CTCGCATAGT 1740
 GGCAGCAGCA GCGGTGGCAG TAAACACAGT GCCGACGGAA TACCACCCAC TGTTCTGAGG 1800
 AGTCCTGTTG GCCTGAGCAG TGATGGCATT TCCTCTAGCT CCAGCTCTTC AAGGAAGAGG 1860
 CTGCATGTCA ATGATGCATC TCACAACCAC CACTCCAAAA TGAGCAAAAG TTCCAAAAGT 1920
 TCAGGTAGTT CATCTAGTTC TTCCTCCTCT GTTAAGCAGT ATATATCCTC TCACAACCTCT 1980
 GTTTTTAACCT ACCCTTACCC CCTCCTCCCC TGTCACATAC CAGGTGGGCT ACGGACATCT 2040
 CTGCACCTCG TGAAACTGGA CAAGAAGCCA GTGGAGACCA ACGGTCCCTGA TGCCAATCAC 2100
 GAGTACAGTA CAAGCAGCCA GCATATGGAC TACAAAGACA CATTGACAT GCTGGACTCA 2160
 CTGTTAAGTG CCCAAGGAAT GAACATGTAA 2190

(2) INFORMATION FOR SEQ ID NO:47:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 729 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

Met Ala Ser Gly Arg Gly Ala Ser Ser Arg Trp Phe Phe Thr Arg Glu
1 5 10 15

Gln Leu Glu Asn Thr Pro Ser Arg Arg Cys Gly Val Glu Ala Asp Lys
20 25 30

Glu Leu Ser Cys Arg Gln Gln Ala Ala Asn Leu Ile Gln Glu Met Gly
35 40 45

Gln Arg Leu Asn Val Ser Gln Leu Thr Ile Asn Thr Ala Ile Val Tyr
50 55 60

Met His Arg Phe Tyr Met His His Ser Phe Thr Lys Phe Asn Lys Asn
65 70 75 80

Ile Ile Ser Ser Thr Ala Leu Phe Leu Ala Ala Lys Val Glu Glu Gln
85 90 95

Ala Arg Lys Leu Glu His Val Ile Lys Val Ala His Ala Cys Leu His
100 105 110

Pro Leu Glu Pro Leu Leu Asp Thr Lys Cys Asp Ala Tyr Leu Gln Gln
115 120 125

Thr Gln Glu Leu Val Ile Leu Glu Thr Ile Met Leu Gln Thr Leu Gly
130 135 140

Phe Glu Ile Thr Ile Glu His Pro His Thr Asp Val Val Lys Cys Thr
145 150 155 160

Gln Leu Val Arg Ala Ser Lys Asp Leu Ala Gln Thr Ser Tyr Phe Met
165 170 175

Ala Thr Asn Ser Leu His Leu Thr Thr Phe Cys Leu Gln Tyr Lys Pro
180 185 190

Thr Val Ile Ala Cys Val Cys Ile His Leu Ala Cys Lys Trp Ser Asn
195 200 205

Trp Glu Ile Pro Val Ser Thr Asp Gly Lys His Trp Trp Glu Tyr Val
210 215 220

Asp Pro Thr Val Thr Leu Glu Leu Leu Asp Glu Leu Thr His Glu Phe
225 230 235 240

Leu Gln Ile Leu Glu Lys Thr Pro Asn Arg Leu Lys Lys Ile Arg Asn
245 250 255

Trp Arg Ala Asn Gln Ala Ala Arg Lys Pro Lys Val Asp Gly Gln Val
260 265 270

Ser Glu Thr Pro Leu Leu Gly Ser Ser Leu Val Gln Asn Ser Ile Leu
275 280 285

Val Asp Ser Val Thr Gly Val Pro Thr Asn Pro Ser Phe Gln Lys Pro
290 295 300

Ser Thr Ser Ala Phe Pro Ala Pro Val Pro Leu Asn Ser Gly Asn Ile
305 310 315 320

Ser Val Gln Asp Ser His Thr Ser Asp Asn Leu Ser Met Leu Ala Thr
325 330 335

Gly Met Pro Ser Thr Ser Tyr Gly Leu Ser Ser His Gln Glu Trp Pro
340 345 350

Gln His Gln Asp Ser Ala Arg Thr Glu Gln Leu Tyr Ser Gln Lys Gln
355 360 365

Glu Thr Ser Leu Ser Gly Ser Gln Tyr Asn Ile Asn Phe Gln Gln Gly
370 375 380

Pro Ser Ile Ser Leu His Ser Gly Leu His His Arg Pro Asp Lys Ile
385 390 395 400

Ser Asp His Ser Ser Val Lys Gln Glu Tyr Thr His Lys Ala Gly Ser
405 410 415

Ser Lys His His Gly Pro Ile Ser Thr Thr Pro Gly Ile Ile Pro Gln
420 425 430

Lys Met Ser Leu Asp Lys Tyr Arg Glu Lys Arg Lys Leu Glu Thr Leu
435 440 445

Asp Leu Asp Val Arg Asp His Tyr Ile Ala Ala Gln Val Glu Gln Gln
450 455 460

His Lys Gln Gly Gln Ser Gln Ala Ala Ser Ser Ser Val Thr Ser
465 470 475 480

Pro Ile Lys Met Lys Ile Pro Ile Ala Asn Thr Glu Lys Tyr Met Ala
485 490 495

Asp Lys Lys Glu Lys Ser Gly Ser Leu Lys Leu Arg Ile Pro Ile Pro
500 505 510

Pro Thr Asp Lys Ser Ala Ser Lys Glu Glu Leu Lys Met Lys Ile Lys
515 520 525

Val Ser Ser Ser Glu Arg His Ser Ser Ser Asp Glu Gly Ser Gly Lys
530 535 540

Ser Lys His Ser Ser Pro His Ile Ser Arg Asp His Lys Glu Lys His
545 550 555 560

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2360 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GGAAGTGCCT GCAACCTTCG CCGCTGCCCTT CTGGTTGAAG CACTATGGAG GGAGAGAGGA 60
AGAACAAACAA CAAACGGTGG TATTCACTC GAGAACAGCT GGAAAATAGC CCATCCCGTC 120
GTTTGCGGT GGACCCAGAT AAAGAACATT CTTATGCCA GCAGGGGCC AATCTGCTTC 180
AGGACATGGG GCAGCGTCTT AACGTCTCAC AATTGACTAT CAACACTGCT ATAGTATAACA 240
TGCATCGATT CTACATGATT CAGTCCTTCA CACGGTTCCC TGGAAATTCT GTGGCTCCAG 300

CAGCCTTGT	TCTAGCAGCT	AAAGTGGAGG	AGCAGCCAA	AAAATTGAA	CATGTCATCA	360
AGGTAGCACA	TACTTGTCTC	CATCCTCAGG	AATCCCTTC	TGATACTAGA	AGTGAGGCTT	420
ATTGCAACA	AGTTCAAGAT	CTGGTCATT	TAGAAAGCAT	AATTTGCAG	ACTTTAGGCT	480
TTGAACTAAC	AATTGATCAC	CCACATACTC	ATGTAGTAAA	GTGCACTCAA	CTTGGTCGAG	540
CAAGCAAGGA	CTTAGCACAG	ACTTCTTACT	TCATGGCAAC	CAACAGCCTG	CATTTGACCA	600
CATTTAGCCT	GCAGTACACA	CCTCCTGTGG	TGGCCTGTGT	CTGCATTCAC	CTGGCTTGCA	660
AGTGGTCAA	TTGGGAGATC	CCAGTCTCAA	CTGACGGAA	GCACTGGTGG	GAGTATGTTG	720
ACGCCACTGT	GACCTTGGAA	CTTTTAGATG	AACTGACACA	TGAGTTCTA	CAGATTTGG	780
AGAAAAACTCC	CAACAGGCTC	AAACGCATT	GGAATTGGAG	GGCATGCGAG	GCTGCCAAGA	840
AAACAAAAGC	AGATGACCGA	GGAACAGATG	AAAAGACTTC	AGAGCAGACA	ATCCTCAATA	900
TGATTTCCA	GAGCTCTTCA	GACACAACCA	TTGCAGGTTT	AATGAGCATG	TCAACTTCTA	960
CCACAAGTGC	AGTGCCTTCC	CTGCCAGTCT	CCGAAGAGTC	ATCCAGCAAC	TTAACCCAGTG	1020
TGGAGATGTT	GCCGGGCAAG	CGTTGGCTGT	CCTCCCAACC	TTCTTTCAA	CTAGAACCTA	1080
CTCAGGGTCA	TCGGACTAGT	GAGAATTAG	CACTTACAGG	AGTTGATCAT	TCCTTACAC	1140
AGGATGGTTC	AAATGCATT	ATTTCCCAGA	AGCAGAATAG	TAAGAGTGTG	CCATCAGCTA	1200
AAGTGTCACT	GAAAGAATAC	CGCGCGAAGC	ATGCAGAAGA	ATTGGCTGCC	CAGAAGAGGC	1260
AACTGGAGAA	CATGGAAGCC	AATGTGAAGT	CACAATATGC	ATATGCTGCC	CAGAATCTCC	1320
TTTCTCATCA	TGATAGCCAT	TCTTCAGTCA	TTCTAAAAAT	GCCCATAGAG	GGTCAGAAA	1380
ACCCCGAGCG	GCCTTTCTG	GAAAAGGCTG	ACAAAACAGC	TCTAAAATG	AGAATCCCAG	1440
TGGCAGGTGG	AGATAAAGCT	GCGTCTTCAA	AACCAGAGGA	GATAAAAATG	CGCATAAAAG	1500
TCCATGCTGC	AGCTGATAAG	CACAATTCTG	TAGAGGACAG	TGTTACAAAG	AGCCGAGAGC	1560
ACAAAGAAGA	GCGCAAGACT	CACCCATCTA	ATCATCATCA	TCATCATAAT	CACCACTCAC	1620
ACAAGCACTC	TCATTCCAA	CTTCCAGTTG	GTACTGGAA	CAAACGTCT	GGTGATCCAA	1680
AACATAGTAG	CCAGACAAGC	AACTTAGCAC	ATAAAACCTA	TAGCTTGTCT	AGTTCTTTT	1740
CCTCTTCCAG	TTCTACTCGT	AAAAGGGGAC	CCTCTGAAGA	GACTGGAGGG	GCTGTGTTG	1800
ATCATCCAGC	CAAGATTGCC	AAGAGTACTA	AATCCTCTTC	CCTAAATTTC	TCCTTCCCTT	1860
CACTTCCTAC	AATGGGTCAG	ATGCCTGGGC	ATAGCTCAGA	CACAAGTGGC	CTTTCCTTTT	1920

CACAGCCCCAG	CTGTAAAACT	CGTGTCCCTC	ATTGAAACT	GGATAAAGGG	CCCACTGGGG	1980
CCAATGGTCA	CAACACGACC	CAGACAATAG	ACTATCAAGA	CACTGTGAAT	ATGCTTCACT	2040
CCCTGCTCAG	TGCCAGGGT	GTTCAAGCCC	CTCAGCCCAC	TGCATTTGAA	TTTGGTCGTC	2100
CTTATAGTGA	CTATCTGAAT	CCTCGGTCTG	GTGGAATCTC	CTCGAGATCT	GGCAATACAG	2160
ACAAACCCCCG	GCCACCACCT	CTGCCATCAG	AACCTCCTCC	ACCACTTCCA	CCCCTTCCTA	2220
AGTAAAAAAA	GAAAAGAAG	AGGAGAAAAA	AACTTCTTTA	AAAAAACACA	TAATTTTCT	2280
TTTTTTTTTG	GGGAAAAAAA	AATTTTTTTT	AAAATTTTTT	CCCCAAGGGA	CGGGGGAAAA	2340
TTTTATTTTT	AAAATTTTTT					2360

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2181 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

ATGGAGGGAG	AGAGGAAGAA	CAACAACAAA	CGGTGGTATT	TCACTCGAGA	ACAGCTGGAA	60
AATAGCCCAT	CCCGTCGTTT	TGGCGTGGAC	CCAGATAAAG	AACTTTCTTA	TCGCCAGCAG	120
GCGGCCAATC	TGCTTCAGGA	CATGGGGCAG	CGTCTTAACG	TCTCACAAATT	GACTATCAAC	180
ACTGCTATAG	TATACATGCA	TCGATTCTAC	ATGATTCACT	CCTTCACACG	GTTCCCTGGA	240
AATTCTGTGG	CTCCAGCAGC	CTTGTCTCTA	GCAGCTAAAG	TGGAGGAGCA	GCCCCAAAAAA	300
TTGGAACATG	TCATCAAGGT	AGCACATACT	TGTCTCCATC	CTCAGGAATC	CCTTCCTGAT	360
ACTAGAACATG	AGGCTTATTT	GCAACAAGTT	CAAGATCTGG	TCATTTAGA	AAGCATAATT	420
TTGCAGACTT	TAGGCTTTGA	ACTAACAAATT	GATCACCCAC	ATACTCATGT	AGTAAAGTGC	480
ACTCAACTTG	TTCGAGCAAG	CAAGGACTTA	GCACAGACTT	CTTACTTCAT	GGCAACCAAC	540
AGCCTGCATT	TGACCACATT	TAGCCTGCAG	TACACACCTC	CTGTGGTGGC	CTGTGTCTGC	600
ATTCACCTGG	CTTGCAAGTG	GTCCAATTGG	GAGATCCCAG	TCTCAACTGA	CGGGAAAGCAC	660
TGGTGGGAGT	ATGTTGACGC	CACTGTGACC	TTGGAACTTT	TAGATGAACT	GACACATGAG	720
TTTCTACAGA	TTTGGAGAA	AACTCCCAAC	AGGCTCAAAC	GCATTTGGAA	TTGGAGGGCA	780

TGCGAGGCTG	CCAAGAAAAC	AAAAGCAGAT	GACCGAGGAA	CAGATGAAAA	GACTTCAGAG	840
CAGACAATCC	TCAATATGAT	TTCCCAGAGC	TCTTCAGACA	CAACCATTGC	AGGTTTAATG	900
AGCATGTCAA	CTTCTACAC	AAGTGCAGTG	CCTTCCCTGC	CAGTCTCCGA	AGAGTCATCC	960
AGCAACTTAA	CCAGTGTGGA	GATGTTGCCG	GGCAAGCGTT	GGCTGTCCTC	CCAACCTTCT	1020
TTCAAACTAG	AACCTACTCA	GGGTCACTCGG	ACTAGTGAGA	ATTTAGCACT	TACAGGAGTT	1080
GATCATT CCT	TACCACAGGA	TGGTTCAAAT	GCATTTATTT	CCCAGAAGCA	GAATAGTAAG	1140
AGTGTGCCAT	CAGCTAAAGT	GTCACTGAAA	GAATACCGCG	CGAACGCATGC	AGAAGAATTG	1200
GCTGCCAGA	AGAGGCAACT	GGAGAACATG	GAAGCCAATG	TGAAGTCACA	ATATGCATAT	1260
GCTGCCAGA	ATCTCCTTTC	TCATCATGAT	AGCCATTCTT	CAGTCATTCT	AAAAATGCC	1320
ATAGAGGGTT	CAGAAAACCC	CGAGCGGCCT	TTTCTGGAAA	AGGCTGACAA	AACAGCTCTC	1380
AAAATGAGAA	TCCCAGTGGC	AGGTGGAGAT	AAAGCTGCGT	CTTCAAAACC	AGAGGAGATA	1440
AAAATGCGCA	TAAAAGTCCA	TGCTGCAGCT	GATAAGCACA	ATTCTGTAGA	GGACAGTGTT	1500
ACAAAGAGCC	GAGAGCACAA	AGAAGAGCGC	AAGACTCACC	CATCTAATCA	TCATCATCAT	1560
CATAATCACC	ACTCACACAA	GCACTCTCAT	TCCCAACTTC	CAGTTGGTAC	TGGGAACAAA	1620
CGTCCTGGTG	ATCCAAAACA	TAGTAGCCAG	ACAAGCAACT	TAGCACATAA	AACCTATAGC	1680
TTGTCTAGTT	CTTTTCCTC	TTCCAGTTCT	ACTCGTAAAA	GGGGACCCTC	TGAAGAGACT	1740
GGAGGGGCTG	TGTTTGATCA	TCCAGCCAAG	ATTGCCAAGA	GTACTAAATC	CTCTCCCTA	1800
AATTTCCTCCT	TCCCTTCACT	TCCTACAATG	GGTCAGATGC	CTGGGCATAG	CTCAGACACA	1860
AGTGGCCTTT	CCTTTTCACA	GCCCAGCTGT	AAAACTCGTG	TCCCTCATTC	GAAACTGGAT	1920
AAAGGGCCCA	CTGGGGCCAA	TGGTCACAAC	ACGACCCAGA	CAATAGACTA	TCAAGACACT	1980
GTGAATATGC	TTCACTCCCT	GCTCAGTGCC	CAGGGTGTTC	AGCCCACCTCA	GCCCAC TGCA	2040
TTTGAATTTG	TTCGTCCTTA	TAGTGACTAT	CTGAATCCTC	GGTCTGGTGG	AATCTCCTCG	2100
AGATCTGGCA	ATACAGACAA	ACCCCGGCCA	CCACCTCTGC	CATCAGAAC	TCCTCCACCA	2160
CTTCCACCCCC	TTCTTAAGTA	A				2181

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 726 amino acids
- (B) TYPE: amino acid

(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

Met Glu Gly Glu Arg Lys Asn Asn Asn Lys Arg Trp Tyr Phe Thr Arg
1 5 10 15

Glu Gln Leu Glu Asn Ser Pro Ser Arg Arg Phe Gly Val Asp Pro Asp
20 25 30

Lys Glu Leu Ser Tyr Arg Gln Gln Ala Ala Asn Leu Leu Gln Asp Met
35 40 45

Gly Gln Arg Leu Asn Val Ser Gln Leu Thr Ile Asn Thr Ala Ile Val
50 55 60

Tyr Met His Arg Phe Tyr Met Ile Gln Ser Phe Thr Arg Phe Pro Gly
65 70 75 80

Asn Ser Val Ala Pro Ala Ala Leu Phe Leu Ala Ala Lys Val Glu Glu
85 90 95

Gln Pro Lys Lys Leu Glu His Val Ile Lys Val Ala His Thr Cys Leu
100 105 110

His Pro Gln Glu Ser Leu Pro Asp Thr Arg Ser Glu Ala Tyr Leu Gln
115 120 125

Gln Val Gln Asp Leu Val Ile Leu Glu Ser Ile Ile Leu Gln Thr Leu
130 135 140

Gly Phe Glu Leu Thr Ile Asp His Pro His Thr His Val Val Lys Cys
145 150 155 160

Thr Gln Leu Val Arg Ala Ser Lys Asp Leu Ala Gln Thr Ser Tyr Phe
165 170 175

Met Ala Thr Asn Ser Leu His Leu Thr Thr Phe Ser Leu Gln Tyr Thr
180 185 190

Pro Pro Val Val Ala Cys Val Cys Ile His Leu Ala Cys Lys Trp Ser
195 200 205

Asn Trp Glu Ile Pro Val Ser Thr Asp Gly Lys His Trp Trp Glu Tyr
210 215 220

Val Asp Ala Thr Val Thr Leu Glu Leu Leu Asp Glu Leu Thr His Glu
225 230 235 240

Phe Leu Gln Ile Leu Glu Lys Thr Pro Asn Arg Leu Lys Arg Ile Trp
245 250 255

DRAFT - NOT FOR DISTRIBUTION

Asn Trp Arg Ala Cys Glu Ala Ala Lys Lys Thr Lys Ala Asp Asp Arg
260 265 270

Gly Thr Asp Glu Lys Thr Ser Glu Gln Thr Ile Leu Asn Met Ile Ser
275 280 285

Gln Ser Ser Ser Asp Thr Thr Ile Ala Gly Leu Met Ser Met Ser Thr
290 295 300

Ser Thr Thr Ser Ala Val Pro Ser Leu Pro Val Ser Glu Glu Ser Ser
305 310 315 320

Ser Asn Leu Thr Ser Val Glu Met Leu Pro Gly Lys Arg Trp Leu Ser
325 330 335

Ser Gln Pro Ser Phe Lys Leu Glu Pro Thr Gln Gly His Arg Thr Ser
340 345 350

Glu Asn Leu Ala Leu Thr Gly Val Asp His Ser Leu Pro Gln Asp Gly
355 360 365

Ser Asn Ala Phe Ile Ser Gln Lys Gln Asn Ser Lys Ser Val Pro Ser
370 375 380

Ala Lys Val Ser Leu Lys Glu Tyr Arg Ala Lys His Ala Glu Glu Leu
385 390 395 400

Ala Ala Gln Lys Arg Gln Leu Glu Asn Met Glu Ala Asn Val Lys Ser
405 410 415

Gln Tyr Ala Tyr Ala Ala Gln Asn Leu Leu Ser His His Asp Ser His
420 425 430

Ser Ser Val Ile Leu Lys Met Pro Ile Glu Gly Ser Glu Asn Pro Glu
435 440 445

Arg Pro Phe Leu Glu Lys Ala Asp Lys Thr Ala Leu Lys Met Arg Ile
450 455 460

Pro Val Ala Gly Gly Asp Lys Ala Ala Ser Ser Lys Pro Glu Glu Ile
465 470 475 480

Lys Met Arg Ile Lys Val His Ala Ala Asp Lys His Asn Ser Val
485 490 495

Glu Asp Ser Val Thr Lys Ser Arg Glu His Lys Glu Glu Arg Lys Thr
500 505 510

His Pro Ser Asn His His His His Asn His His Ser His Lys His
515 520 525

Ser His Ser Gln Leu Pro Val Gly Thr Gly Asn Lys Arg Pro Gly Asp
530 535 540

Pro Lys His Ser Ser Gln Thr Ser Asn Leu Ala His Lys Thr Tyr Ser
 545 550 555 560
 Leu Ser Ser Ser Phe Ser Ser Ser Ser Thr Arg Lys Arg Gly Pro
 565 570 575
 Ser Glu Glu Thr Gly Gly Ala Val Phe Asp His Pro Ala Lys Ile Ala
 580 585 590
 Lys Ser Thr Lys Ser Ser Ser Leu Asn Phe Ser Phe Pro Ser Leu Pro
 595 600 605
 Thr Met Gly Gln Met Pro Gly His Ser Ser Asp Thr Ser Gly Leu Ser
 610 615 620
 Phe Ser Gln Pro Ser Cys Lys Thr Arg Val Pro His Ser Lys Leu Asp
 625 630 635 640
 Lys Gly Pro Thr Gly Ala Asn Gly His Asn Thr Thr Gln Thr Ile Asp
 645 650 655
 Tyr Gln Asp Thr Val Asn Met Leu His Ser Leu Leu Ser Ala Gln Gly
 660 665 670
 Val Gln Pro Thr Gln Pro Thr Ala Phe Glu Phe Val Arg Pro Tyr Ser
 675 680 685
 Asp Tyr Leu Asn Pro Arg Ser Gly Gly Ile Ser Ser Arg Ser Gly Asn
 690 695 700
 Thr Asp Lys Pro Arg Pro Pro Pro Leu Pro Ser Glu Pro Pro Pro Pro
 705 710 715 720
 Leu Pro Pro Leu Pro Lys
 725

(2) INFORMATION FOR SEQ ID NO:51:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

TTCCCCACCAA TGCTTTCC

18

(2) INFORMATION FOR SEQ ID NO:52:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs

- (B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

CCATCAGTTG ATACAGGGAT CT

22

(2) INFORMATION FOR SEQ ID NO:53:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

GGAATTCA GA AGGTTGTAAG ATGC

24

(2) INFORMATION FOR SEQ ID NO:54:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

ACACACAGAT GTGGTGAAAT GTACCCA

27

(2) INFORMATION FOR SEQ ID NO:55:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

GCATCTTACA ACCTTCTG

18

(2) INFORMATION FOR SEQ ID NO:56:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

GGAATTCATG GAAAGCATTG GTGGGAAT

28

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

CCTCCACTAC TGGTTGCCT GG

22

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

GGACTAGTAT AAATATGGCG TCAGGCCGTG

30

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

GGAGATCTTA CATGTTCAATT CCTTGGG

27

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

GGAGACAAGT ATGTGCTACC TTGATGACA

29

(2) INFORMATION FOR SEQ ID NO:61:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

GGAATTCCGGG CTGCTCCTCC ACTTTAG

27

(2) INFORMATION FOR SEQ ID NO:62:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

GGAATTTCGCT GCTGGAGCCA CAGAA

25

(2) INFORMATION FOR SEQ ID NO:63:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

GTGTCACTGA AAGAATAACCG

20

(2) INFORMATION FOR SEQ ID NO:64:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

GGAATTCAAGG TGGAGATAAA GCTGC

25

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

GCTCTAGATA AATATGGAGG GAGAGAGGAA

30

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

GGAATTCTTA CTTAGGAAGG GGTGGAAGTG

30

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

GGAATTCTTA CTTAGGAAGG GGTGGAAGTG GTGGAGGAGG TTAC

44

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

Ala Cys Ser Tyr Ser Pro Thr Ser Pro Ser Tyr Ser Pro Thr Ser Pro
1 5 10 15

Ser Tyr Ser Pro Thr Ser Pro Ser Lys Lys
20 25

WHAT IS CLAIMED IS:

1. A DNA segment comprising an isolated coding region that encodes a substantially full length P-TEFb subunit, wherein the coding region is characterized as:

5

(a) encoding a substantially full length P-TEFb kinase subunit having the amino acid sequence of SEQ ID NO:2; or

10

(b) encoding a substantially full length P-TEFb large subunit that includes a contiguous sequence of at least about 7 amino acids from SEQ ID NO:4, SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50; or as a substantially full length coding region that hybridizes to the nucleotide sequence of SEQ ID NO:3, SEQ ID NO:43 or SEQ ID NO:48 under stringent hybridization conditions.

15

2. The DNA segment of claim 1, wherein said isolated coding region encodes a substantially full length P-TEFb kinase subunit having the amino acid sequence of SEQ ID NO:2.

20

3. The DNA segment of claim 2, wherein said isolated coding region has the nucleotide sequence from position 115 to position 1327 of SEQ ID NO:1.

30

4. The DNA segment of claim 1, wherein said isolated coding region encodes a substantially full length P-TEFb large subunit that includes a contiguous sequence of at least about 7 amino acids from SEQ ID NO:4, SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50; or is a substantially full length coding region that hybridizes to the nucleotide sequence of SEQ ID NO:3, SEQ ID NO:43 or SEQ ID NO:48 under stringent hybridization conditions.

5. The DNA segment of claim 1, wherein said isolated coding region encodes a substantially full length P-TEFb large subunit that includes a contiguous sequence of at least about 7 amino acids from SEQ ID NO:4, SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50 and wherein said coding region hybridizes to the nucleotide sequence of SEQ ID NO:3, SEQ ID NO:43 or
5 SEQ ID NO:48 under stringent hybridization conditions.

6. The DNA segment of claim 4, wherein said isolated coding region encodes a substantially full length P-TEFb large subunit that includes a contiguous sequence of at least about 7 amino acids from SEQ ID NO:4, SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50.
10

7. The DNA segment of claim 6, wherein said isolated coding region encodes a P-TEFb large subunit having the amino acid sequence of SEQ ID NO:4, SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50.
15

8. The DNA segment of claim 7, wherein said isolated coding region encodes a P-TEFb large subunit having the amino acid sequence of SEQ ID NO:45.
20

9. The DNA segment of claim 7, wherein said isolated coding region encodes a P-TEFb large subunit having the amino acid sequence of SEQ ID NO:47.
25

10. The DNA segment of claim 7, wherein said isolated coding region encodes a P-TEFb large subunit having the amino acid sequence of SEQ ID NO:50.

11. The DNA segment of claim 4, wherein said isolated coding region is a substantially full length coding region that hybridizes to the nucleotide sequence of SEQ ID NO:3, SEQ ID NO:43 or SEQ ID NO:48 under stringent hybridization conditions.

5

12. The DNA segment of claim 11, wherein said isolated coding region has the nucleotide sequence of SEQ ID NO:44.

10 13. The DNA segment of claim 11, wherein said isolated coding region has the nucleotide sequence of SEQ ID NO:46.

PCT/GB2008/000489

15 14. The DNA segment of claim 11, wherein said isolated coding region has the nucleotide sequence of SEQ ID NO:49.

20 15. The DNA segment of claim 1, wherein said DNA segment comprises a first coding region that encodes a substantially full length P-TEFb kinase subunit and a second coding region that encodes a substantially full length P-TEFb large subunit.

25 16. The DNA segment of claim 15, wherein said second coding region encodes a P-TEFb large subunit that has the amino acid sequence of SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50.

17. The DNA segment of claim 16, wherein said first coding region encodes a P-TEFb kinase subunit that has the amino acid sequence of SEQ ID NO:6.

30

18. The DNA segment of claim 16, wherein said second coding region has the nucleotide sequence of SEQ ID NO:44, SEQ ID NO:46 or SEQ ID NO:49, and wherein said first coding region has the nucleotide sequence of SEQ ID NO:5.

5

19. The DNA segment of claim 1, wherein said isolated coding region is operatively attached to a second coding region that encodes a selected peptide or protein sequence, said DNA segment encoding a P-TEFb subunit fusion protein.

10

20. The DNA segment of claim 1, operatively positioned under the control of a promoter.

21. The DNA segment of claim 20, further defined as a recombinant vector.

22. The DNA segment of claim 20, comprised within a recombinant host cell.

23. An expression system comprising:

(a) a first expression unit comprising, under the transcriptional control of a promoter, a first coding region that encodes a substantially full length P-TEFb kinase subunit that includes a contiguous sequence of at least about 7 amino acids from SEQ ID NO:2 or SEQ ID NO:6; and

(b) a second expression unit comprising, under the transcriptional control of a promoter, a second coding region that encodes a substantially full length P-TEFb large subunit that includes a contiguous sequence of at least about 7 amino acids from SEQ ID NO:4, SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50.

5

24. The expression system of claim 23, wherein said first and said second expression units are comprised on a single expression vector.

10

25. The expression system of claim 23, wherein said first and said second expression units are comprised on two distinct expression vectors.

卷之三

26. The expression system of claim 23, wherein said expression system is comprised within a recombinant host cell.

27. A recombinant host cell comprising at least a first DNA segment in accordance with claim 1.

28. The recombinant host cell of claim 27, wherein said cell is a prokaryotic host cell

25

29. The recombinant host cell of claim 27, wherein said cell is a eukaryotic host cell

30. The recombinant host cell of claim 27, wherein said cell further comprises an HIV Tat
30 protein.

31. The recombinant host cell of claim 27, wherein said cell comprises a first DNA segment that encodes a substantially full length P-TEFb kinase subunit and a second DNA segment that
5 encodes a substantially full length P-TEFb large subunit.

32. The recombinant host cell of claim 31, wherein said first and second DNA segments are comprised within a single expression vector.

10

33. A method for detecting P-TEFb nucleic acids in a sample, comprising obtaining sample nucleic acids from a sample suspected of containing P-TEFb nucleic acids; contacting said sample nucleic acids with a nucleic acid segment that hybridizes to the sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:43 or SEQ ID NO:48, under conditions effective to allow hybridization of substantially complementary nucleic acids; and detecting the hybridized complementary nucleic acids thus formed.
15

20 34. The method of claim 33, wherein the sample nucleic acids are obtained from a sample suspected of containing a tumor cell.

25 35. A method of using a DNA segment that encodes a substantially full length P-TEFb subunit, comprising expressing at least a first DNA segment in accordance with claim 1 in a recombinant host cell and collecting the P-TEFb subunit expressed by said cell.

36. A composition comprising at least a first isolated, substantially full length P-TEFb protein subunit that includes a contiguous sequence of at least about 7 amino acids from SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50.

5

37. The composition of claim 36, wherein said substantially full length P-TEFb subunit is operatively attached to a selected peptide or protein sequence.

10 38. The composition of claim 36, wherein said composition comprises a substantially full length P-TEFb kinase subunit in operative association with a substantially full length P-TEFb large subunit.

39. The composition of claim 36, further comprising an HIV Tat protein.

40. An isolated, functional P-TEFb enzyme complex comprising a P-TEFb kinase subunit in operative association with a P-TEFb large subunit.

41. A P-TEFb immunodetection reagent characterized as:

(a) an antibody that has immunospecificity for a P-TEFb subunit that includes a contiguous sequence of at least about 7 amino acids from SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:45, SEQ ID NO:47 or SEQ ID NO:50; or

(b) an antibody that has immunospecificity for a P-TEFb subunit that includes a contiguous sequence of at least about 7 amino acids from SEQ ID NO:6, the antibody operatively attached to a detectable label.

42. A method of identifying a gene that encodes a protein that interacts with a P-TEFb subunit, comprising the steps of:

5

- (a) obtaining a first DNA segment comprising a candidate gene; said first DNA segment expressing a first fusion protein comprising a transcriptional transactivating domain operatively attached to the candidate protein encoded by said candidate gene;
- (b) obtaining a second DNA segment that expresses a second fusion protein comprising a P-TEFb subunit operatively attached to a DNA binding domain that binds to a defined nucleic acid sequence;
- (c) providing said first and second DNA segments to a eukaryotic host cell that comprises a marker gene operatively positioned downstream of said defined nucleic acid sequence; and
- (d) identifying a eukaryotic host cell that expresses said marker gene, thereby identifying said candidate gene as a gene that encodes a protein that interacts with a P-TEFb subunit.

15
20

43. A method for identifying a candidate transcriptional inhibitor, comprising preparing a P-TEFb composition comprising at least a P-TEFb kinase subunit and testing said candidate inhibitor for the ability to inhibit P-TEFb-mediated phosphorylation of RNA polymerase II, wherein inhibition of phosphorylation is indicative of a candidate transcriptional inhibitor.

30 44. The method of claim 43, comprising the steps of:

- 5
- (a) obtaining a P-TEFb composition comprising at least a P-TEFb kinase subunit;
 - (b) obtaining an RNA polymerase II composition comprising at least the carboxyl terminal domain (CTD) of the large subunit of RNA polymerase II;
 - 10
 - (c) admixing said P-TEFb composition with said RNA polymerase II composition and an effective phosphate donor compound comprising a labeled phosphate group; and
 - (d) determining the ability of said P-TEFb composition to transfer said labeled phosphate group to said RNA polymerase II composition in the presence of said candidate transcriptional inhibitor and in the absence of said candidate transcriptional inhibitor, wherein a reduction in the amount of labeled phosphate transferred to RNA polymerase II in the presence of said candidate is indicative of a candidate transcriptional inhibitor.
- 15

20 45. The method of claim 43, wherein said P-TEFb composition comprises a P-TEFb enzyme complex that has transcription elongation promoting activity.

25 46. The method of claim 45, further comprising testing the candidate transcriptional inhibitor so identified in a transcription elongation assay, wherein inhibition of transcription elongation confirms the identification of a transcriptional inhibitor.

30 47. The method of claim 43, further comprising the step of purifying the candidate transcriptional inhibitor so identified.

48. A method for identifying an HIV Tat protein, comprising contacting a composition suspected of containing an HIV Tat protein with a human P-TEFb composition under conditions effective to allow the formation of bound protein complexes and detecting the bound Tat:P-TEFb
5 protein complexes so formed.

49. A method for identifying a candidate viral transcription inhibitor, comprising testing a candidate substance for the ability to inhibit the binding of a viral transcriptional transactivator
10 protein to at least one human P-TEFb subunit under effective binding conditions, wherein inhibition of binding is indicative of a candidate viral transcription inhibitor.

15 50. The method of claim 49, wherein said viral transcriptional transactivator protein or said at least one human P-TEFb subunit is attached to a solid support.

20 51. The method of claim 49, wherein said effective binding conditions comprise admixing a human cell nuclear extract with said viral transcriptional transactivator protein and said at least one human P-TEFb subunit.

25 52. The method of claim 49, wherein said inhibition of binding is tested by determining the ability of the candidate substance to inhibit the formation of viral protein-P-TEFb subunit complexes that are detected by means of a detectable label attached to the viral protein or to the P-TEFb subunit.

53. The method of claim 49, wherein said inhibition of binding is tested by determining the ability of the candidate substance to inhibit the formation of viral protein-P-TEFb subunit complexes that are detected by means of a first specific immunological detection reagent.

5

54. The method of claim 53, wherein said inhibition of binding is tested by determining the ability of the candidate substance to inhibit the formation of viral protein-P-TEFb subunit complexes that are detected by means of a first and a second specific immunological detection reagent.

10

55. The method of claim 49, further comprising testing the candidate viral transcription inhibitor so identified in a viral transcription elongation assay, wherein inhibition of viral transcription elongation confirms the identity of an active candidate viral transcription inhibitor.

P0001
P0002
P0003
P0004
P0005
P0006
P0007
P0008
P0009
P0010
P0011
P0012
P0013
P0014
P0015
P0016
P0017
P0018
P0019
P0020

56. The method of claim 55, further comprising testing the active candidate viral transcription inhibitor next identified in separate human and viral transcription elongation assays, wherein inhibition of viral transcription elongation, and not human transcription elongation, in the presence of said viral transcriptional transactivator protein confirms the identification of an active viral transcription inhibitor.

25 57. The method of claim 49, further comprising the step of purifying the candidate viral transcriptional inhibitor so identified.

30

58. The method of claim 57, wherein said purified candidate viral transcriptional inhibitor is formulated in a pharmaceutically acceptable vehicle.

59. A method for identifying a candidate viral transcription inhibitor, comprising testing a candidate substance for the ability to inhibit viral RNA elongation in a functional viral transcription elongation assay, wherein inhibition of viral RNA elongation is indicative of an
5 active candidate viral transcription inhibitor.

60. The method of claim 59, further comprising testing the active candidate viral transcription inhibitor in distinct human and viral transcription elongation assays, wherein the
10 inhibition of viral, but not human, transcription elongation confirms the identity of an active candidate viral transcription inhibitor.

61. The method of claim 60, wherein said method comprises the steps of:

- (a) preparing a first transcriptionally competent composition capable of generating elongated human RNA transcripts, said composition comprising effective amounts of human nucleic acid template, P-TEFb enzyme complex, RNA polymerase II, the required nucleotide triphosphates and ATP;
- (b) preparing a second transcriptionally competent composition capable of generating elongated viral RNA transcripts, said composition comprising effective amounts of viral nucleic acid template, viral transcriptional transactivator protein, P-TEFb enzyme complex, RNA polymerase II, the required nucleotide triphosphates and ATP; and
- (c) identifying a viral transcription inhibitor that inhibits the generation of elongated viral RNA transcripts by said second transcriptionally competent composition but that does not inhibit the generation of elongated human RNA transcripts by said first transcriptionally competent composition.

62. The method of claim 61, wherein said candidate viral transcription inhibitor is a candidate HIV transcription inhibitor and wherein said method comprises the steps of:

5

- (a) preparing a first transcriptionally competent composition capable of generating elongated human RNA transcripts, said composition comprising effective amounts of human nucleic acid template, P-TEFb enzyme complex, RNA polymerase II, the required nucleotide triphosphates and ATP;
- (b) preparing a second transcriptionally competent composition capable of generating elongated HIV RNA transcripts, said composition comprising effective amounts of HIV nucleic acid template, HIV Tat protein, P-TEFb enzyme complex, RNA polymerase II, the required nucleotide triphosphates and ATP; and
- (c) identifying an HIV inhibitor that inhibits the generation of elongated HIV RNA transcripts by said second transcriptionally competent composition but that does not inhibit the generation of elongated human RNA transcripts by said first transcriptionally competent composition.

10

15

20

63. The method of claim 59, further comprising the step of purifying the candidate viral transcriptional inhibitor so identified.

25

64. The method of claim 63, wherein said purified candidate viral transcriptional inhibitor is formulated in a pharmaceutically acceptable vehicle.

65. A method for inhibiting viral replication, comprising contacting a cell suspected of being infected with a virus with a biologically effective amount of a viral transcription inhibitor identified by the method of claim 59.

5

66. The method of claim 65, wherein said cell is a cell suspected of being infected with HIV.

67. The method of claim 65, wherein said cell is located within an animal and a
10 therapeutically effective amount of said inhibitor is administered to said animal.

2025 RELEASE UNDER E.O. 14176

ABSTRACT

Disclosed is the discovery that the transcription elongation factor termed P-TEFb has a central role in transcription elongation control. P-TEFb is herein shown to phosphorylate RNA polymerase II and to control the transition from abortive into productive elongation mode. P-TEFb has also been discovered to interact with the HIV transcriptional transactivating protein, Tat, showing that P-TEFb is the cellular factor necessary for HIV Tat to effect productive viral mRNA elongation. The invention provides genes encoding P-TEFb subunits, including human genes, and related biological components, and also provides assay methods connected with the control of transcription elongation. Particularly useful assays are those concerning the identification of substances that inhibit viral replication at the transcription elongation stage by inhibiting the binding or functional interaction of viral proteins to P-TEFb.

FIG. 1

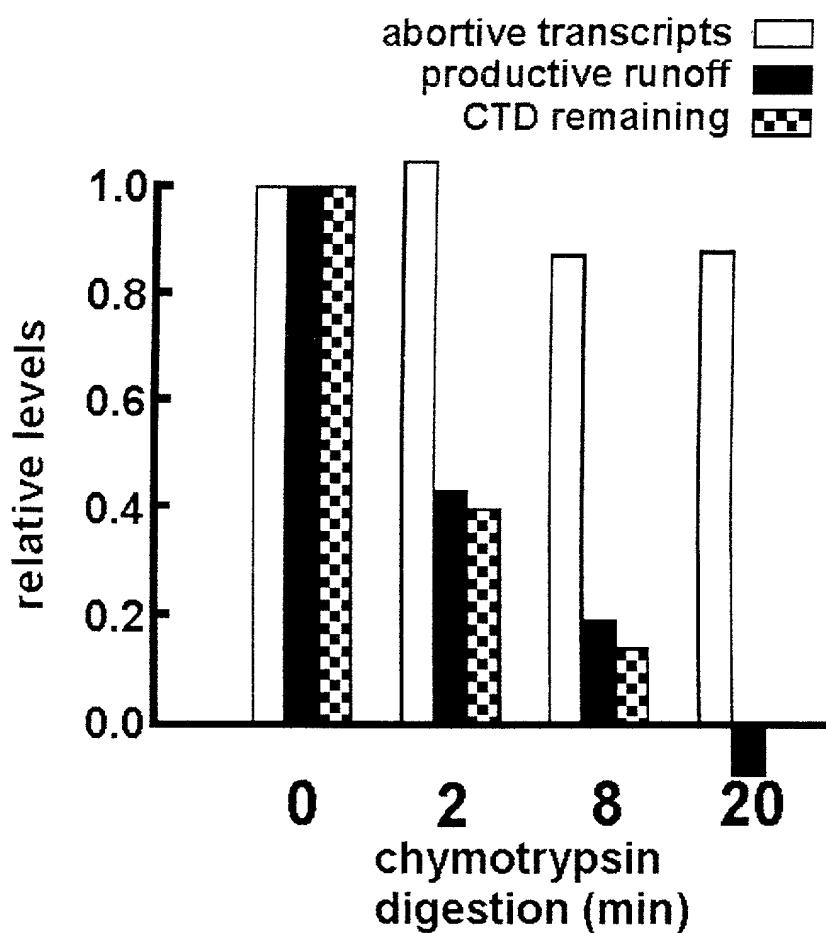


FIG. 2A

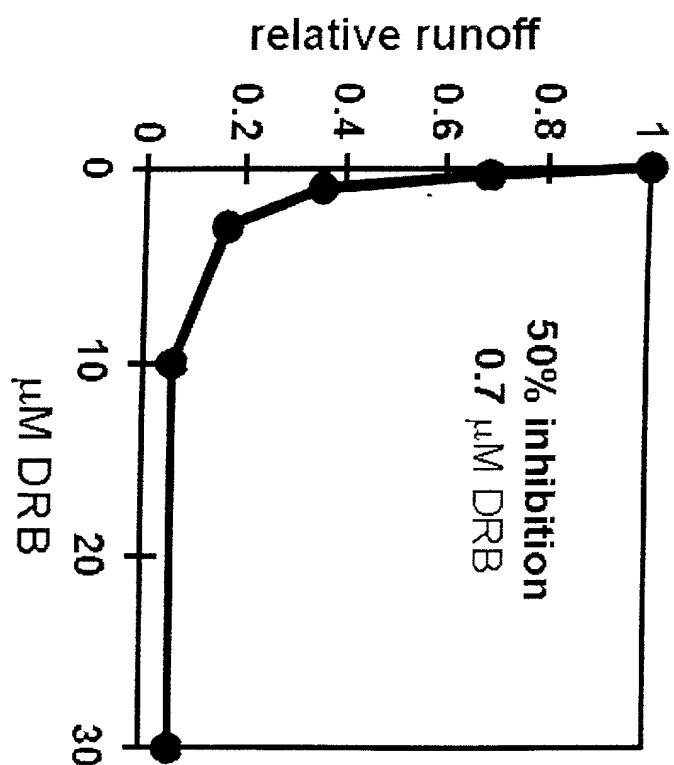


FIG. 2B

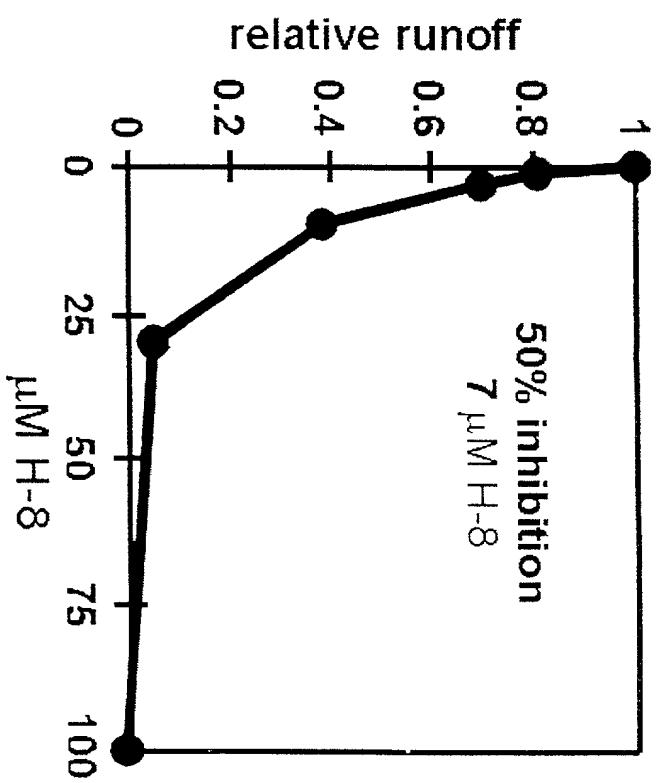


FIG. 3

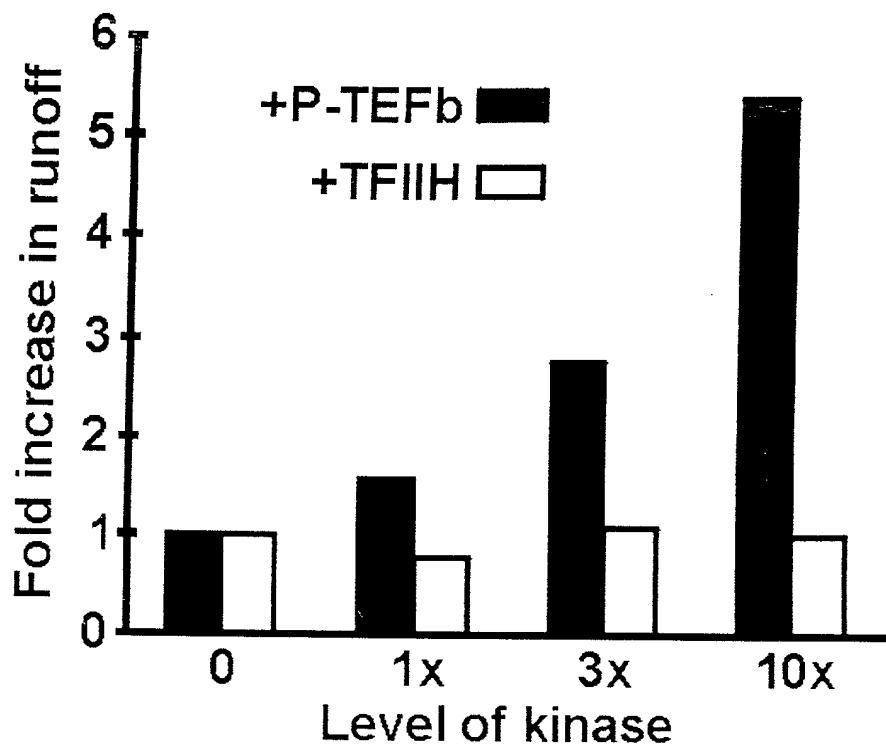


FIG. 4

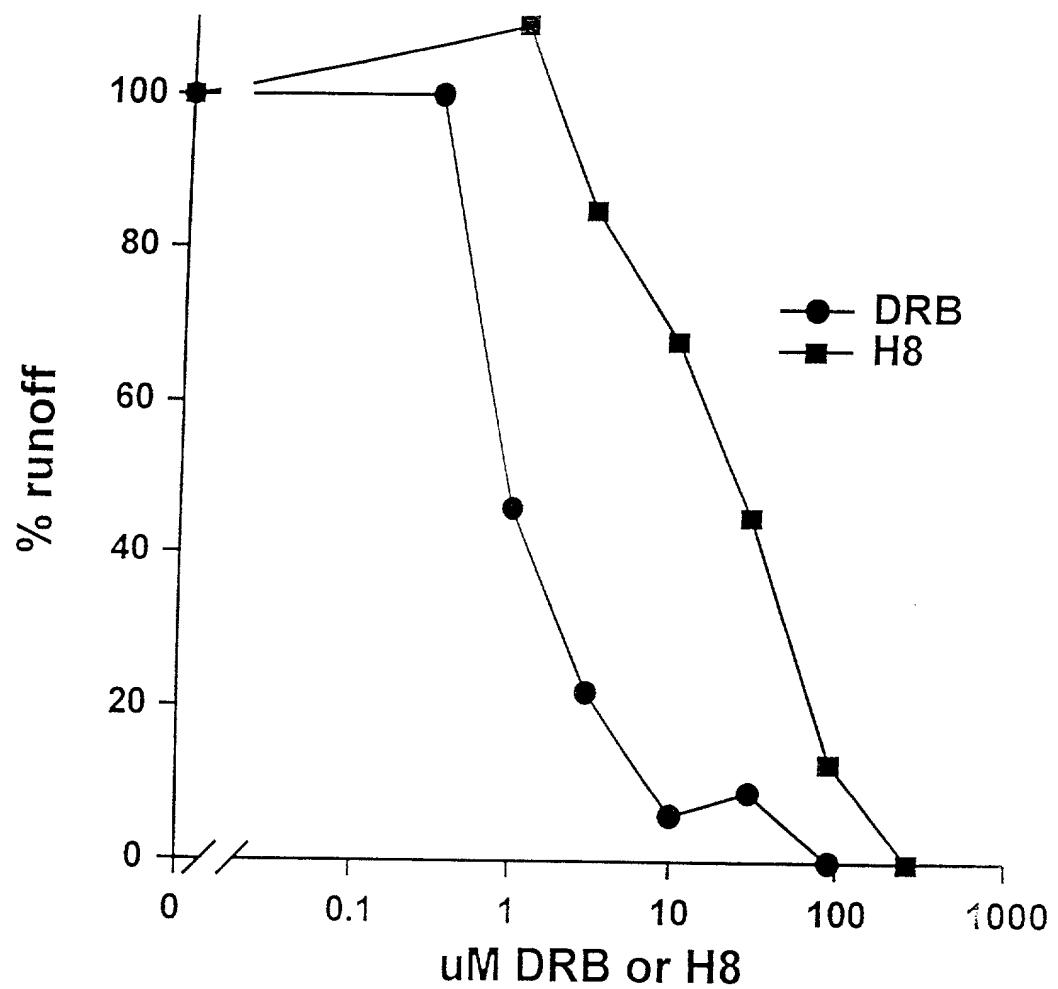
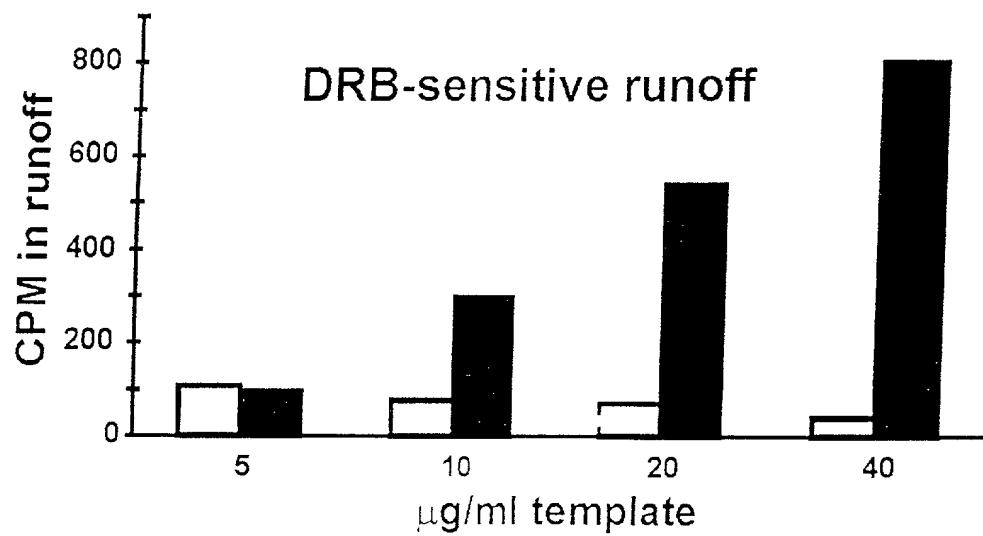


FIG. 5



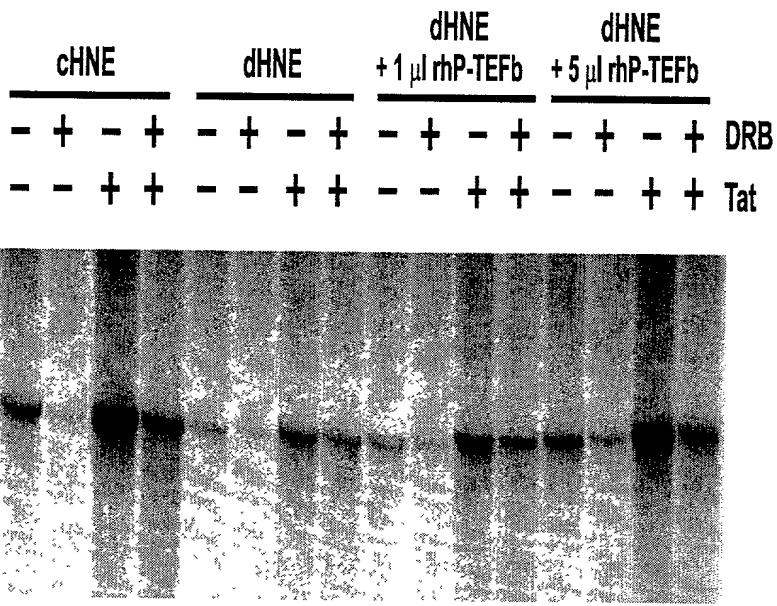


FIG. 6

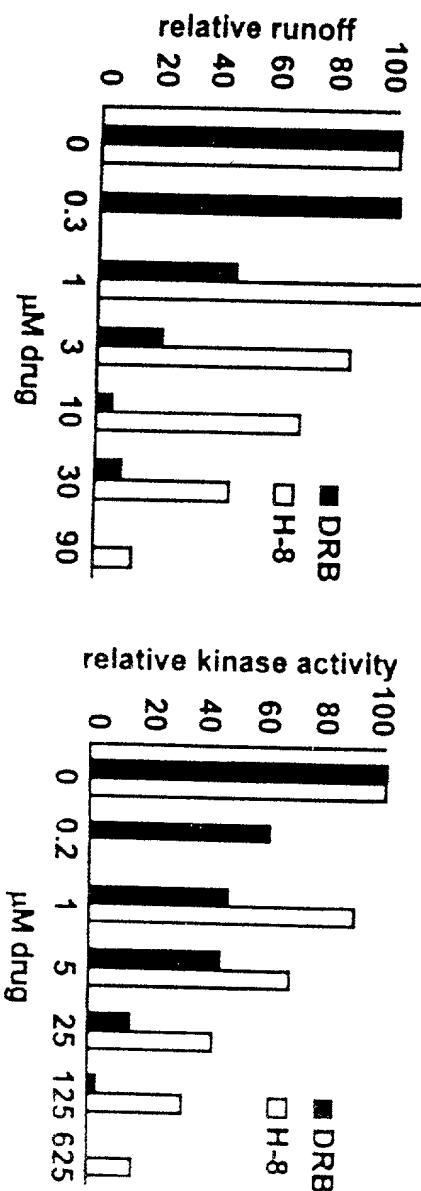


FIG. 7A

FIG. 7B

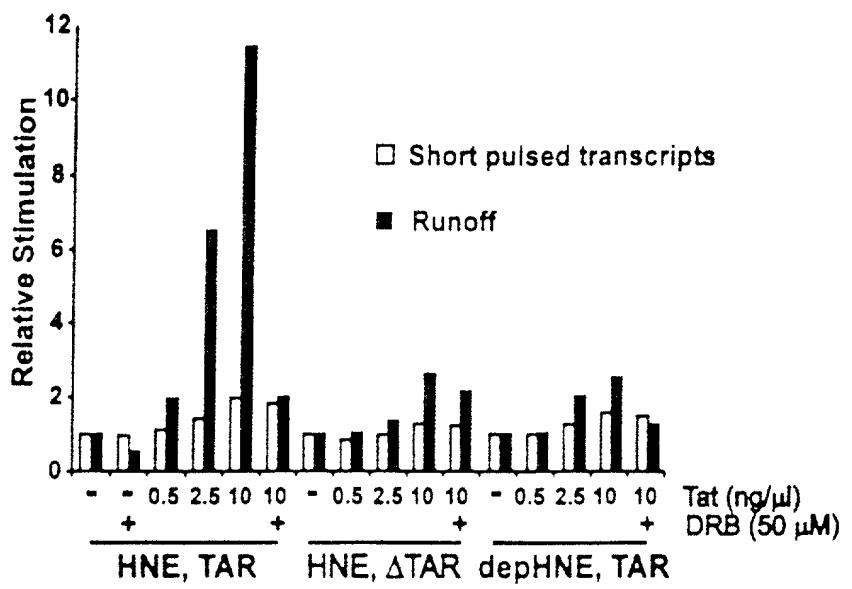


FIG. 8

SEQUENCE LISTING

70647 U.S. PTO
08/951188
10/15/97



(1) GENERAL INFORMATION:

(i) APPLICANT: Price, David H.

(ii) TITLE OF INVENTION: P-TEF^b COMPOSITIONS, METHODS AND SCREENING ASSAYS

(iii) NUMBER OF SEQUENCES: 68

(iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: Arnold, White & Durkee
- (B) STREET: P.O. Box 4433
- (C) CITY: Houston
- (D) STATE: TX
- (E) COUNTRY: USA
- (F) ZIP: 77210-4433

(v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER: US Unknown
- (B) FILING DATE: Concurrently Herewith
- (C) CLASSIFICATION: Unknown

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Fussey, Shelley P.M.
- (B) REGISTRATION NUMBER: 39,458
- (C) REFERENCE/DOCKET NUMBER: IOWA:012

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: (512) 418-3000
- (B) TELEFAX: (512) 418-3131

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1457 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 115..1326

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TGTTGAGTCA ACAGCTGTAG ATACACCAAT TGTTGCCGAT TTCTTTCTTT TCGACTGTG	60
GCTTCTCGCG AAACTGTGAT TGTGAAAATT GTACAAATAG AGGCAAATTT AACC ATG	117
Met	
1	
GCG CAC ATG TCC CAC ATG CTC CAG CAG CCT TCG GGG TCG ACG CCC TCC	165
Ala His Met Ser His Met Leu Gln Gln Pro Ser Gly Ser Thr Pro Ser	
5 10 15	
AAC GTG GGC TCC AGC TCA TCG CGC ACG ATG TCC CTG ATG GAG AAA CAA	213
Asn Val Gly Ser Ser Ser Arg Thr Met Ser Leu Met Glu Lys Gln	
20 25 30	
AAG TAC ATC GAG GAC TAC GAC TTT CCC TAC TGC GAC GAG AGC AAC AAA	261
Lys Tyr Ile Glu Asp Tyr Asp Phe Pro Tyr Cys Asp Glu Ser Asn Lys	
35 40 45	
TAC GAA AAG GTG GCG AAA ATT GGC CAA GGC ACC TTC GGA GAG GTT TTT	309
Tyr Glu Lys Val Ala Lys Ile Gly Gln Gly Thr Phe Gly Glu Val Phe	
50 55 60 65	
AAG GCT CGC GAG AAA AAG GGC AAC AAG AAG TTT GTG GCC ATG AAG AAG	357
Lys Ala Arg Glu Lys Lys Gly Asn Lys Lys Phe Val Ala Met Lys Lys	
70 75 80	
GTG CTG ATG GAC AAC GAA AAG GAG GGC TTT CCC ATC ACG GCT CTG CGA	405
Val Leu Met Asp Asn Glu Lys Glu Gly Phe Pro Ile Thr Ala Leu Arg	
85 90 95	
GAG ATC CGC ATC CTG CAG CTG CTA AAG CAC GAG AAC GTG GTG AAT CTG	453
Glu Ile Arg Ile Leu Gln Leu Lys His Glu Asn Val Val Asn Leu	
100 105 110	
ATC GAG ATC TGC CGC ACC AAG GCC ACC GCC ACG AAT GGT TAC AGA TCC	501
Ile Glu Ile Cys Arg Thr Lys Ala Thr Ala Thr Asn Gly Tyr Arg Ser	
115 120 125	
ACC TTC TAT TTG GTC TTT GAT TTC TGC GAA CAC GAT TTG GCA GGT CTT	549
Thr Phe Tyr Leu Val Phe Asp Phe Cys Glu His Asp Leu Ala Gly Leu	
130 135 140 145	
CTG TCC AAC ATG AAC GTC AAG TTC AGT CTG GGC GAG ATT AAG AAG GTT	597
Leu Ser Asn Met Asn Val Lys Phe Ser Leu Gly Glu Ile Lys Lys Val	
150 155 160	
ATG CAG CAG CTT TTA AAC GGT TTG TAT TAC ATC CAC AGC AAC AAG ATC	645
Met Gln Gln Leu Leu Asn Gly Leu Tyr Tyr Ile His Ser Asn Lys Ile	
165 170 175	
CTG CAC CGA GAC ATG AAA GCT GCC AAC GTG CTG ATT ACC AAG CAT GGC	693
Leu His Arg Asp Met Lys Ala Ala Asn Val Leu Ile Thr Lys His Gly	
180 185 190	

DRAFT

ATC TTA AAG CTG GCT GAC TTT GGC TTG GCC CGT GCT TTT AGC ATT CCA Ile Leu Lys Leu Ala Asp Phe Gly Leu Ala Arg Ala Phe Ser Ile Pro 195 200 205	741
AAG AAC GAG AGT AAG AAT CGC TAT ACC AAT CGC GTA GTA ACC TTG TGG Lys Asn Glu Ser Lys Asn Arg Tyr Thr Asn Arg Val Val Thr Leu Trp 210 215 220 225	789
TAC CGG CCG CCT GAG CTG CTA CTT GGT GAC CGC AAC TAT GGT CCA CCC Tyr Arg Pro Pro Glu Leu Leu Gly Asp Arg Asn Tyr Gly Pro Pro 230 235 240	837
GTG GAC ATG TGG GGA GCC GGC TGC ATA ATG GCC GAG ATG TGG ACA CGC Val Asp Met Trp Gly Ala Gly Cys Ile Met Ala Glu Met Trp Thr Arg 245 250 255	885
TCG CCC ATC ATG CAA GGC AAT ACG GAG CAG CAG CAG TTA ACC TTT ATT Ser Pro Ile Met Gln Gly Asn Thr Glu Gln Gln Leu Thr Phe Ile 260 265 270	933
TCG CAG CTA TGC GGC TCC TTT ACG CCG GAC GTG TGG CCG GGA GTG GAG Ser Gln Leu Cys Gly Ser Phe Thr Pro Asp Val Trp Pro Gly Val Glu 275 280 285	981
GAG CTG GAG CTG TAC AAA TCC ATC GAG CTG CCA AAG AAC CAG AAG CGT Glu Leu Glu Leu Tyr Lys Ser Ile Glu Leu Pro Lys Asn Gln Lys Arg 290 295 300 305	1029
CGA GTC AAG GAG CGC CTG CGT CCG TAT GTC AAG GAT CAA ACC GGC TGT Arg Val Lys Glu Arg Leu Arg Pro Tyr Val Lys Asp Gln Thr Gly Cys 310 315 320	1077
GAT CTA TTG GAC AAA TTG CTG ACC CTT GAT CCC AAG AAA CGC ATC GAT Asp Leu Leu Asp Lys Leu Leu Thr Leu Asp Pro Lys Lys Arg Ile Asp 325 330 335	1125
GCG GAC ACA GCT CTG AAT CAC GAC TTC TTC TGG ACG GAT CCC ATG CCC Ala Asp Thr Ala Leu Asn His Asp Phe Phe Trp Thr Asp Pro Met Pro 340 345 350	1173
AGC GAC TTG AGC AAG ATG CTG TCC CAG CAC CTG CAG AGC ATG TTC GAG Ser Asp Leu Ser Lys Met Leu Ser Gln His Leu Gln Ser Met Phe Glu 355 360 365	1221
TAC CTG GCG CAG CCA CGC CGC AGC AAC CAG ATG CGC AAC TAT CAC CAG Tyr Leu Ala Gln Pro Arg Arg Ser Asn Gln Met Arg Asn Tyr His Gln 370 375 380 385	1269
CAA CTG ACC ACC ATG AAC CAG AAG CCC CAG GAC AAC AGT ATG ATT GAC Gln Leu Thr Thr Met Asn Gln Lys Pro Gln Asp Asn Ser Met Ile Asp 390 395 400	1317
CGG GTT TGG TAGACTGCCA GAGGTGTACG CACCCGACTA ATAGTTCTC	1366

Arg Val Trp

ACCTTCAACT AGCGTTAGGT TATTAGGTTA GTGTACAATA AAAATATTGG CATTTCGATT 1426
AGCGCTTGCT CCAAATATAA AAAAAAAAAA A 1457

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:

 - (A) LENGTH: 404 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ala His Met Ser His Met Leu Gln Gln Pro Ser Gly Ser Thr Pro			
1	5	10	15
Ser Asn Val Gly Ser Ser Ser Arg Thr Met Ser Leu Met Glu Lys			
20	25	30	
Gln Lys Tyr Ile Glu Asp Tyr Asp Phe Pro Tyr Cys Asp Glu Ser Asn			
35	40	45	
Lys Tyr Glu Lys Val Ala Lys Ile Gly Gln Gly Thr Phe Gly Glu Val			
50	55	60	
Phe Lys Ala Arg Glu Lys Lys Gly Asn Lys Lys Phe Val Ala Met Lys			
65	70	75	80
Lys Val Leu Met Asp Asn Glu Lys Glu Gly Phe Pro Ile Thr Ala Leu			
85	90	95	
Arg Glu Ile Arg Ile Leu Gln Leu Leu Lys His Glu Asn Val Val Asn			
100	105	110	
Leu Ile Glu Ile Cys Arg Thr Lys Ala Thr Ala Thr Asn Gly Tyr Arg			
115	120	125	
Ser Thr Phe Tyr Leu Val Phe Asp Phe Cys Glu His Asp Leu Ala Gly			
130	135	140	
Leu Leu Ser Asn Met Asn Val Lys Phe Ser Leu Gly Glu Ile Lys Lys			
145	150	155	160
Val Met Gln Gln Leu Leu Asn Gly Leu Tyr Tyr Ile His Ser Asn Lys			
165	170	175	
Ile Leu His Arg Asp Met Lys Ala Ala Asn Val Leu Ile Thr Lys His			
180	185	190	

Gly Ile Leu Lys Leu Ala Asp Phe Gly Leu Ala Arg Ala Phe Ser Ile
195 200 205

Pro Lys Asn Glu Ser Lys Asn Arg Tyr Thr Asn Arg Val Val Thr Leu
210 215 220

Trp Tyr Arg Pro Pro Glu Leu Leu Leu Gly Asp Arg Asn Tyr Gly Pro
225 230 235 240

Pro Val Asp Met Trp Gly Ala Gly Cys Ile Met Ala Glu Met Trp Thr
245 250 255

Arg Ser Pro Ile Met Gln Gly Asn Thr Glu Gln Gln Gln Leu Thr Phe
260 265 270

Ile Ser Gln Leu Cys Gly Ser Phe Thr Pro Asp Val Trp Pro Gly Val
275 280 285

Glu Glu Leu Glu Leu Tyr Lys Ser Ile Glu Leu Pro Lys Asn Gln Lys
290 295 300

Arg Arg Val Lys Glu Arg Leu Arg Pro Tyr Val Lys Asp Gln Thr Gly
305 310 315 320

Cys Asp Leu Leu Asp Lys Leu Leu Thr Leu Asp Pro Lys Lys Arg Ile
325 330 335

Asp Ala Asp Thr Ala Leu Asn His Asp Phe Phe Trp Thr Asp Pro Met
340 345 350

Pro Ser Asp Leu Ser Lys Met Leu Ser Gln His Leu Gln Ser Met Phe
355 360 365

Glu Tyr Leu Ala Gln Pro Arg Arg Ser Asn Gln Met Arg Asn Tyr His
370 375 380

Gln Gln Leu Thr Thr Met Asn Gln Lys Pro Gln Asp Asn Ser Met Ile
385 390 395 400

Asp Arg Val Trp

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4328 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

P05371-20040709-2

CAGCCCTGCC GACGGCCATA CTTGAAAATA CATTTCCTTC TGCAAAGTTT GTCATTGTCA 60
CTGTGTGAAT GGAATCTGTG ATGTGTTGTG GAATTAAAAA CGTCAAGTAA ACAACCCGTA 120
ATGGTTAAAG TGCACGGCGA AAGCAGTGC G AATAACTATG AATTGATACA AAAGTTGCAT 180
AACACGTCGC CTGGTGTGCG GGTTAGTGTG TTTTCGTCT CGTTTCGTTT CCGCCGCAGT 240
CGCAGTTCC AAAAACCTC ACCACACCAC ACCATCTCCA CCACGCACAC ACACACACAA 300
ACAAACACGC AGAGACGCGG CGGCGGAAAA AGTGTGCGGA CCGCGGATT AACCCTCGT 360
TCCAAACCCA AATTGGAGTC TCCCACAAAC AGCGAAATAT CGAGTGTGGC TTAGCCGATG 420
TGCCGTGCGA TCCCCACTGC CCCTTCCGTA CCGCTGCCAC CCCC GCCACA GCAGCAACGC 480
ACACGGATAC GGACACAGAC ACCAATACCA GCGCACTCAA GCACGGCCGA CAAAGAAAGA 540
GCGCTCTCCC TTCCTCTTG TACAGTTAGT TCCTACAGCT GAATCAGCCA AAAGAAATTA 600
CTAGGTCCAT TCCGAGGCGC AGTTGCATG TGAAACGGAG GTCCCCGCAT AACCAACGCGG 660
AACCCGAAAT TCCAGATCCC CATCTCCGCT GCACGGATAA AGGAAACATA CAACCATGAG 720
TCTCCTAGCC ACGCCAATGC CCCAGGCGC CACCGCCTCA TCTTCTTCAT CCGCCTCCGC 780
GGCCGCCTCG GCCAGCGGGG TTCCAATCAC CGCCAACAAAC AACCTGCCTT TCGAGAAGGA 840
CAAGATCTGG TACTTCAGCA ACGATCAGCT GGCCAATTG CCAAGCAGAA GATGCGGCAT 900
CAAGGGCGAC GATGAGCTGC AGTACCGCCA GATGACCGCC TATCTGATAC AGGAAATGGG 960
TCAGCGTCTG CAGGTGTCCC AACTGTGCAT CAACACGGCC ATTGTGTACA TGCATCGGTT 1020
CTACGCCTTT CACTCCTTCA CCCACTTCA TCGCAACTCC ATGGCGTCGG CGAGCCTCTT 1080
CTTGGCCGCC AAGGTAGAAG AGCAACCGCG GAAGCTGGAG CATGTTATTC GGGCCGCCAA 1140
CAAGTGCCTG CCGCCGACCA CCGAGCAGAA TTACGCCGAA CTCGCCAGG AGCTTGTGTT 1200
CAACGAGAAC GTGCTCCTGC AGACGCTGGG CTTCGATGTG GCCATCGATC ATCCGCACAC 1260
GCATGTGGTG CGCACCTGCC AGCTGGTCAA AGCATGCAAG GATCTGGCGC AGACATCGTA 1320
CTTCTTGGCC TCGAACAGCC TGCATCTGAC CTCGATGTG CTCCAATATC GCCCCACGGT 1380
CGTAGCCTGT TTCTGCATTT ACCTAGCCTG CAAGTGGTCC CGATGGGAGA TCCCCCAGTC 1440
GACCGAGGGC AAGCACTGGT TCTACTATGT GGACAAGACG GTCTCGCTGG ATTTGCTAAA 1500
GCAGCTGACA GATGAGTTCA TCGCTATCTA TGAGAAGAGC CCGGCCCCGTC TGAAGTCTAA 1560
GCTTAACTCG ATCAAGGCGA TCGCCCCAGGG AGCCAGCAAT CGGACAGCTA ACAGCAAGGA 1620

CAAACCAAAG GAGGACTGGA AGATCACCGA GATGATGAAG GGCTACCACT CAAACATCAC 1680
GACACCACCA GAGCTGTTAA ACGGCAACGA CAGCCGGGAT CGGGACCGAG ATCGTGAACG 1740
GGAGAGAGAG CGGGAACGGG ATCCGTCGTC ACTACTGCCG CCACCGGCTA TGGTGCCGCA 1800
GCAAAGACGA CAGGATGGTG GACATCAGCG CTCGTCCCTCA GTGAGCGGAG TGCCAGGCAG 1860
CAGCTCTTCG TCGTCTTCCT CCAGTCACAA GATGCCAAAT TACCCCTGGTG GCATGCCGCC 1920
CGAAGCTCAT CCGGATCACA AGTCAAAGCA GCCGGGCTAT AACAAATCGAA TGCCCTCAAG 1980
TCACCAGCGT AGTAGTAGCA GTGGACTCGG TTCCTCGGGA AGTGGCAGCC AGCACAGCAG 2040
CTCATCCTCG TCGTCTTCAA GCCAGCAGCC TGGCCGACCG TCTATGCCCG TGGACTATCA 2100
CAAATCCTCT CGCGGCATGC CGCCGGTAGG CGTGGGCATG CCACCTCACG GCAGCCACAA 2160
GATGACTTCG GGCTCCAAGC CTCAACAGCC GCAGCAGCAG CCGGTCCCAC ATCCATCCGC 2220
CTCTAATTCC TCTGCATCGG GCATGTCTC CAAGGATAAA TCCCAGAGCA ACAAAATGTA 2280
TCCGAACGCA CCGCCGCCAT ACAGTAATAG TGCCCCCTCAA AACCCGCTGA TGTGGCGTGG 2340
TGGATATCCA GGCCTAGCA ATGGATCCCA GCCCCCGCCT CCCGCCGGAT ACGGCGGCCA 2400
TCGCAGCAAA TCCGGCTCCA CCGTCCATGG CATGCCGCAT TTGAGCAGC AATTGCCCTA 2460
TTCCCAGAGC CAGAGCTACG GCCACATGCA GCAGCAGCCA GTGCCTCAGT CTCAGCAGCA 2520
ACAGATGCCT CCGGAGGCAT CCCAGCACTC GTTGCAGTCC AAGAACTCGC TCTTCAGTCC 2580
AGAGTGGCCA GACATTAAAA AGGAGCCCAT GTCGCAGTCG CAACCACAGC TTTTTAACGG 2640
TTTGCTACCC CCTCCTGCAC CTCCCGGCCA CGATTACAAG CTAAATAGCC ATCCGCGCGA 2700
CAAAGAAAGT CCCAAGAAAG AGCGACTAAC GCCAACCAAA AAGGATAAGC ACCGTCTGT 2760
AATGCCCTCA ATGGGCAGTG GGAACAGTTC CTCCGGCTCG GGATCATCAA AGCCGATGCT 2820
ACCGCCTCAC AAGAACGAGA TACCCATGG CGGGGACCTG TTGACCAATC CTGGAGAGAG 2880
TGGAAGCCTA AAACGGCCA ACGAGATCTC GGGAAAGTCAG TATGGACTAA ATAAGCTGGA 2940
TGAAATAGAT AACAGTAATA TGCCTCGAGA AAAGCTTCGC AAGCTGGACA CTACAACGG 3000
ACTACCAACT TATCCGAATT ATGAGGAGAA ACACACGCCT CTGAATATGT CCAACGGAAT 3060
CGAGACAACG CCGGATCTGG TGCGCAGTTT GCTAAAGGAG AGTCTGTGTC CATCGAACGC 3120
TTCGCTCCTG AAACCGGATG CCTTGACTAT GCCTGGCCTG AAACCAACGG CGAACTACT 3180
TGAGCCCAGT CCCGCACCAAG CGACAATCAA GAAAGAACAG GGAATAACTC CGATGACCAG 3240

DRAFT

TTGGCTAGT GGGCCCGCAC CCATGGATT GGAAGTACCC ACTAACAGG CCGGAGAGAT	3300
TAAGGAGGAA AGCAGCAGCA AGTCCGAAAA GAAAAGAAG AAGGATAAAC ACAAACACAA	3360
GGAGAAGGAC AAGTCCAAGG ACAAGACGGA AAAGGAGGAG CGTAAGAACG ACAAGAGGGA	3420
CAAGCAGAAG GATCGTAGCG GCAGCGGTGG CAGCAAGGAC AGTTCTCTTC CCAATGAGCC	3480
TCTGAAGATG GTTATCAAGA ATCCCACGG CAGCCTGCAG GCCGGTGCAGT CAGCTCCCAT	3540
TAAACTTAAG ATCAGCAAAA ATAAGGTTGA ACCCAATAAC TACTCTGCAG CGGCGGGTCT	3600
GCCTGGCGCA ATCGGATATG GCTTGCCCTCC AACTACGGCT ACCACCACAT CCGCTTCGAT	3660
CGGAGCAGCT GCTCCTGTTG TGCCCTCTTA TGGTGCCGGC GGTGGTGGCT ACAGCTCATC	3720
GGCGGGCAGC AGTTCCGGTG GCAGCAGCAA GAAAAGCAC AGCGATCGTG ACCGCGACAA	3780
GGAGAGCAAA AAGAATAAGA GCCAAGACTA CGCGAAGTAC AATGGCGCTG GTGGCGGCAT	3840
CTTTAATCCC CTTGGCGGTG CTGGCGCCGC ACCCAATATG TCTGGAGGAA TGGGCGCCCC	3900
CATGTCTACT GCTGTACAC CATCCATGCT GTTGGCGCCC ACCGGTGCAG TACCACCCCTC	3960
TGCCGCTGGG CTGGCACCGC CTCCCATGCC CGTCTACAAC AAGAAGTAGT GGTAGCGGTC	4020
AGAGGGTTAT TCTTAAGTCG TACGTTTGA TATATGTATA GAACCTCAGT AAGTCCGATT	4080
GTAGTATAGT TGTTAGGATT GTTAGTGAGA TGCATTATTG ATTTAGTTA AGCACATAGA	4140
TAAAACCCA AATTGGAAGT GAAACCGGAT GCGCAGATCG AAGAAGAATG GAAGTAGATG	4200
TCGCGATGGG GCTGGACGTA AAAGCAGTAC TCAAATCGCG AAAACTTTG TACAGCATTA	4260
ATTAGTTAT AACTATAATA AATAGCATAAC ATATAAGCCC AAAAAAAA AAAAAAAA	4320
AAAAAAA	4328

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1097 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Ser Leu Leu Ala Thr Pro Met Pro Gln Ala Ala Thr Ala Ser Ser			
1	5	10	15

Ser Ser Ser Ala Ser Ala Ala Ser Ala Ser Gly Ile Pro Ile Thr		
20	25	30

Ala Asn Asn Asn Leu Pro Phe Glu Lys Asp Lys Ile Trp Tyr Phe Ser
 35 40 45
 Asn Asp Gln Leu Ala Asn Leu Pro Ser Arg Arg Cys Gly Ile Lys Gly
 50 55 60
 Asp Asp Glu Leu Gln Tyr Arg Gln Met Thr Ala Tyr Leu Ile Gln Glu
 65 70 75 80
 Met Gly Gln Arg Leu Gln Val Ser Gln Leu Cys Ile Asn Thr Ala Ile
 85 90 95
 Val Tyr Met His Arg Phe Tyr Ala Phe His Ser Phe Thr His Phe His
 100 105 110
 Arg Asn Ser Met Ala Ser Ala Ser Leu Phe Leu Ala Ala Lys Val Glu
 115 120 125
 Glu Gln Pro Arg Lys Leu Glu His Val Ile Arg Ala Ala Asn Lys Cys
 130 135 140
 Leu Pro Pro Thr Thr Glu Gln Asn Tyr Ala Glu Leu Ala Gln Glu Leu
 145 150 155 160
 Val Phe Asn Glu Asn Val Leu Leu Gln Thr Leu Gly Phe Asp Val Ala
 165 170 175
 Ile Asp His Pro His Thr His Val Val Arg Thr Cys Gln Leu Val Lys
 180 185 190
 Ala Cys Lys Asp Leu Ala Gln Thr Ser Tyr Phe Leu Ala Ser Asn Ser
 195 200 205
 Leu His Leu Thr Ser Met Cys Leu Gln Tyr Arg Pro Thr Val Val Ala
 210 215 220
 Cys Phe Cys Ile Tyr Leu Ala Cys Lys Trp Ser Arg Trp Glu Ile Pro
 225 230 235 240
 Gln Ser Thr Glu Gly Lys His Trp Phe Tyr Tyr Val Asp Lys Thr Val
 245 250 255
 Ser Leu Asp Leu Leu Lys Gln Leu Thr Asp Glu Phe Ile Ala Ile Tyr
 260 265 270
 Glu Lys Ser Pro Ala Arg Leu Lys Ser Lys Leu Asn Ser Ile Lys Ala
 275 280 285
 Ile Ala Gln Gly Ala Ser Asn Arg Thr Ala Asn Ser Lys Asp Lys Pro
 290 295 300
 Lys Glu Asp Trp Lys Ile Thr Glu Met Met Lys Gly Tyr His Ser Asn
 305 310 315 320

Ile Thr Thr Pro Pro Glu Leu Leu Asn Gly Asn Asp Ser Arg Asp Arg
325 330 335

Asp Arg Asp Arg Glu Arg Glu Arg Glu Arg Asp Pro Ser Ser
340 345 350

Leu Leu Pro Pro Ala Met Val Pro Gln Gln Arg Arg Gln Asp Gly
355 360 365

Gly His Gln Arg Ser Ser Ser Val Ser Gly Val Pro Gly Ser Ser Ser
370 375 380

Ser Ser Ser Ser Ser His Lys Met Pro Asn Tyr Pro Gly Gly Met
385 390 395 400

Pro Pro Glu Ala His Pro Asp His Lys Ser Lys Gln Pro Gly Tyr Asn
405 410 415

Asn Arg Met Pro Ser Ser His Gln Arg Ser Ser Ser Ser Gly Leu Gly
420 425 430

Ser Ser Gly Ser Gly Ser Gln His Ser Ser Ser Ser Ser Ser Ser Ser
435 440 445

Ser Gln Gln Pro Gly Arg Pro Ser Met Pro Val Asp Tyr His Lys Ser
450 455 460

Ser Arg Gly Met Pro Pro Val Gly Val Gly Met Pro Pro His Gly Ser
465 470 475 480

His Lys Met Thr Ser Gly Ser Lys Pro Gln Gln Pro Gln Gln Pro
485 490 495

Val Pro His Pro Ser Ala Ser Asn Ser Ser Ala Ser Gly Met Ser Ser
500 505 510

Lys Asp Lys Ser Gln Ser Asn Lys Met Tyr Pro Asn Ala Pro Pro Pro
515 520 525

Tyr Ser Asn Ser Ala Pro Gln Asn Pro Leu Met Ser Arg Gly Gly Tyr
530 535 540

Pro Gly Ala Ser Asn Gly Ser Gln Pro Pro Pro Pro Ala Gly Tyr Gly
545 550 555 560

Gly His Arg Ser Lys Ser Gly Ser Thr Val His Gly Met Pro His Phe
565 570 575

Glu Gln Gln Leu Pro Tyr Ser Gln Ser Gln Ser Tyr Gly His Met Gln
580 585 590

Gln Gln Pro Val Pro Gln Ser Gln Gln Gln Met Pro Pro Glu Ala
595 600 . 605

Ser Gln His Ser Leu Gln Ser Lys Asn Ser Leu Phe Ser Pro Glu Trp
610 615 620

Pro Asp Ile Lys Lys Glu Pro Met Ser Gln Ser Gln Pro Gln Leu Phe
625 630 635 640

Asn Gly Leu Leu Pro Pro Pro Ala Pro Pro Gly His Asp Tyr Lys Leu
645 650 655

Asn Ser His Pro Arg Asp Lys Glu Ser Pro Lys Lys Glu Arg Leu Thr
660 665 670

Pro Thr Lys Lys Asp Lys His Arg Pro Val Met Pro Pro Met Gly Ser
675 680 685

Gly Asn Ser Ser Ser Gly Ser Gly Ser Ser Lys Pro Met Leu Pro Pro
690 695 700

His Lys Lys Gln Ile Pro His Gly Gly Asp Leu Leu Thr Asn Pro Gly
705 710 715 720

Glu Ser Gly Ser Leu Lys Arg Pro Asn Glu Ile Ser Gly Ser Gln Tyr
725 730 735

Gly Leu Asn Lys Leu Asp Glu Ile Asp Asn Ser Asn Met Pro Arg Glu
740 745 750

Lys Leu Arg Lys Leu Asp Thr Thr Thr Gly Leu Pro Thr Tyr Pro Asn
755 760 765

Tyr Glu Glu Lys His Thr Pro Leu Asn Met Ser Asn Gly Ile Glu Thr
770 775 780

Thr Pro Asp Leu Val Arg Ser Leu Leu Lys Glu Ser Leu Cys Pro Ser
785 790 795 800

Asn Ala Ser Leu Leu Lys Pro Asp Ala Leu Thr Met Pro Gly Leu Lys
805 810 815

Pro Pro Ala Glu Leu Leu Glu Pro Met Pro Ala Pro Ala Thr Ile Lys
820 825 830

Lys Glu Gln Gly Ile Thr Pro Met Thr Ser Leu Ala Ser Gly Pro Ala
835 840 845

Pro Met Asp Leu Glu Val Pro Thr Lys Gln Ala Gly Glu Ile Lys Glu
850 855 860

Glu Ser Ser Ser Lys Ser Glu Lys Lys Lys Lys Lys Asp Lys His Lys
865 870 875 880

His Lys Glu Lys Asp Lys Ser Lys Asp Lys Thr Glu Lys Glu Glu Arg
885 890 895

Lys Lys His Lys Arg Asp Lys Gln Lys Asp Arg Ser Gly Ser Gly Gly
 900 905 910

 Ser Lys Asp Ser Ser Leu Pro Asn Glu Pro Leu Lys Met Val Ile Lys
 915 920 925

 Asn Pro Asn Gly Ser Leu Gln Ala Gly Ala Ser Ala Pro Ile Lys Leu
 930 935 940

 Lys Ile Ser Lys Asn Lys Val Glu Pro Asn Asn Tyr Ser Ala Ala Ala
 945 950 955 960

 Gly Leu Pro Gly Ala Ile Gly Tyr Gly Leu Pro Pro Thr Thr Ala Thr
 965 970 975

 Thr Thr Ser Ala Ser Ile Gly Ala Ala Ala Pro Val Leu Pro Pro Tyr
 980 985 990

 Gly Ala Gly Gly Gly Tyr Ser Ser Ser Gly Gly Ser Ser Ser Gly
 995 1000 1005

 Gly Ser Ser Lys Lys His Ser Asp Arg Asp Arg Asp Lys Glu Ser
 1010 1015 1020

 Lys Lys Asn Lys Ser Gln Asp Tyr Ala Lys Tyr Asn Gly Ala Gly Gly
 1025 1030 1035 1040

 Gly Ile Phe Asn Pro Leu Gly Gly Ala Gly Ala Ala Pro Asn Met Ser
 1045 1050 1055

 Gly Gly Met Gly Ala Pro Met Ser Thr Ala Val Pro Pro Ser Met Leu
 1060 1065 1070

 Leu Ala Pro Thr Gly Ala Val Pro Pro Ser Ala Ala Gly Leu Ala Pro
 1075 1080 1085

 Pro Pro Met Pro Val Tyr Asn Lys Lys
 1090 1095

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1119 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..1116

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

PROTEIN SEQUENCES

the first time in the history of the world, the people of the United States have been compelled to make a choice between two political parties.

ATG GCA AAG CAG TAC GAC TCG GTG GAG TGC CCT TTT TGT GAT GAA GTT Met Ala Lys Gln Tyr Asp Ser Val Glu Cys Pro Phe Cys Asp Glu Val	1	5	10	15	48
TCC AAA TAC GAG AAG CTC GCC AAG ATC GGC CAA GGC ACC TTC GGG GAG Ser Lys Tyr Glu Lys Leu Ala Lys Ile Gly Gln Gly Thr Phe Gly Glu	20	25	30		96
GTC TTC AAG GCC AGG CAC CGC AAG ACC GGC CAG AAG GTG GCT CTG AAG Val Phe Lys Ala Arg His Arg Lys Thr Gly Gln Lys Val Ala Leu Lys	35	40	45		144
AAG GTG CTG ATG GAA AAC GAG AAG GAG GGG TTC CCC ATT ACA GCC TTG Lys Val Leu Met Glu Asn Glu Lys Glu Gly Phe Pro Ile Thr Ala Leu	50	55	60		192
CGG GAG ATC AAG ATC CTT CAG CTT CTA AAA CAC GAG AAT GTG GTC AAC Arg Glu Ile Lys Ile Leu Gln Leu Leu Lys His Glu Asn Val Val Asn	65	70	75	80	240
TTG ATT GAG ATT TGT CGA ACC AAA GCT TCC CCC TAT AAC CGC TGC AAG Leu Ile Glu Ile Cys Arg Thr Lys Ala Ser Pro Tyr Asn Arg Cys Lys	85	90	95		288
GGT AGT ATA TAC CTG GTG TTC GAC TTC TGC GAG CAT GAC CTT GCT GGG Gly Ser Ile Tyr Leu Val Phe Asp Phe Cys Glu His Asp Leu Ala Gly	100	105	110		336
CTG TTG AGC AAT GTT TTG GTC AAG TTC ACG CTG TCT GAG ATC AAG AGG Leu Leu Ser Asn Val Leu Val Lys Phe Thr Leu Ser Glu Ile Lys Arg	115	120	125		384
GTG ATG CAG ATG CTG CTT AAC GGC CTC TAC TAC ATC CAC AGA AAC AAG Val Met Gln Met Leu Leu Asn Gly Leu Tyr Tyr Ile His Arg Asn Lys	130	135	140		432
ATC CTG CAT AGG GAC ATG AAG GCT AAT GTG CTT ATC ACT CGT GAT Ile Leu His Arg Asp Met Lys Ala Ala Asn Val Leu Ile Thr Arg Asp	145	150	155	160	480
GGG GTC CTG AAG CTG GCA GAC TTT GGG CTG GCC CGG GCC TTC AGC CTG Gly Val Leu Lys Leu Ala Asp Phe Gly Leu Ala Arg Ala Phe Ser Leu	165	170	175		528
GCC AAG AAC AGC CAG CCC AAC CGC TAC ACC AAC CGT GTG GTG ACA CTC Ala Lys Asn Ser Gln Pro Asn Arg Tyr Thr Asn Arg Val Val Thr Leu	180	185	190		576
TGG TAC CGG CCC CCG GAG CTG TTG CTC GGG GAG CGG GAC TAC GGC CCC Trp Tyr Arg Pro Pro Glu Leu Leu Leu Gly Glu Arg Asp Tyr Gly Pro	195	200	205		624
CCC ATT GAC CTG TGG GGT GCT GGG TGC ATC ATG GCA GAG ATG TGG ACC					672

Pro Ile Asp Leu Trp Gly Ala Gly Cys Ile Met Ala Glu Met Trp Thr
 210 215 220

CGC AGC CCC ATC ATG CAG GGC AAC ACG GAG CAG CAC CAA CTC GCC CTC 720
 Arg Ser Pro Ile Met Gln Gly Asn Thr Glu Gln His Gln Leu Ala Leu
 225 230 235 240

ATC AGT CAG CTC TGC GGC TCC ATC ACC CCT GAG GTG TGG CCA AAC GTG 768
 Ile Ser Gln Leu Cys Gly Ser Ile Thr Pro Glu Val Trp Pro Asn Val
 245 250 255

GAC AAC TAT GAG CTG TAC GAA AAG CTG GAG CTG GTC AAG GGC CAG AAG 816
 Asp Asn Tyr Glu Leu Tyr Glu Lys Leu Glu Leu Val Lys Gly Gln Lys
 260 265 270

CGG AAG GTG AAG GAC AGG CTG AAG GCC TAT GTG CGT GAC CCA TAC GCA 864
 Arg Lys Val Lys Asp Arg Leu Lys Ala Tyr Val Arg Asp Pro Tyr Ala
 275 280 285

CTG GAC CTC ATC GAC AAG CTG CTG GTG CTG GAC CCT GCC CAG CGC ATC 912
 Leu Asp Leu Ile Asp Lys Leu Leu Val Leu Asp Pro Ala Gln Arg Ile
 290 295 300

GAC AGC GAT GAC GCC CTC AAC CAC GAC TTC TTC TGG TCC GAC CCC ATG 960
 Asp Ser Asp Asp Ala Leu Asn His Asp Phe Phe Trp Ser Asp Pro Met
 305 310 315 320

CCC TCC GAC CTC AAG GGC ATG CTC TCC ACC CAC CTG ACG TCC ATG TTC 1008
 Pro Ser Asp Leu Lys Gly Met Leu Ser Thr His Leu Thr Ser Met Phe
 325 330 335

GAG TAC TTG GCA CCA CCG CGC CGG AAG GGC AGC CAG ATC ACC CAG CAG 1056
 Glu Tyr Leu Ala Pro Pro Arg Arg Lys Gly Ser Gln Ile Thr Gln Gln
 340 345 350

TCC ACC AAC CAG AGT CGC AAT CCC GCC ACC ACC AAC CAG ACG GAG TTT 1104
 Ser Thr Asn Gln Ser Arg Asn Pro Ala Thr Thr Asn Gln Thr Glu Phe
 355 360 365

GAG CGC GTC TTC TGA 1119
 Glu Arg Val Phe
 370

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 372 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

DRAFT

Met Ala Lys Gln Tyr Asp Ser Val Glu Cys Pro Phe Cys Asp Glu Val
1 5 10 15

Ser Lys Tyr Glu Lys Leu Ala Lys Ile Gly Gln Gly Thr Phe Gly Glu
20 25 30

Val Phe Lys Ala Arg His Arg Lys Thr Gly Gln Lys Val Ala Leu Lys
35 40 45

Lys Val Leu Met Glu Asn Glu Lys Glu Gly Phe Pro Ile Thr Ala Leu
50 55 60

Arg Glu Ile Lys Ile Leu Gln Leu Leu Lys His Glu Asn Val Val Asn
65 70 75 80

Leu Ile Glu Ile Cys Arg Thr Lys Ala Ser Pro Tyr Asn Arg Cys Lys
85 90 95

Gly Ser Ile Tyr Leu Val Phe Asp Phe Cys Glu His Asp Leu Ala Gly
100 105 110

Leu Leu Ser Asn Val Leu Val Lys Phe Thr Leu Ser Glu Ile Lys Arg
115 120 125

Val Met Gln Met Leu Leu Asn Gly Leu Tyr Tyr Ile His Arg Asn Lys
130 135 140

Ile Leu His Arg Asp Met Lys Ala Ala Asn Val Leu Ile Thr Arg Asp
145 150 155 160

Gly Val Leu Lys Leu Ala Asp Phe Gly Leu Ala Arg Ala Phe Ser Leu
165 170 175

Ala Lys Asn Ser Gln Pro Asn Arg Tyr Thr Asn Arg Val Val Thr Leu
180 185 190

Trp Tyr Arg Pro Pro Glu Leu Leu Gly Glu Arg Asp Tyr Gly Pro
195 200 205

Pro Ile Asp Leu Trp Gly Ala Gly Cys Ile Met Ala Glu Met Trp Thr
210 215 220

Arg Ser Pro Ile Met Gln Gly Asn Thr Glu Gln His Gln Leu Ala Leu
225 230 235 240

Ile Ser Gln Leu Cys Gly Ser Ile Thr Pro Glu Val Trp Pro Asn Val
245 250 255

Asp Asn Tyr Glu Leu Tyr Glu Lys Leu Glu Leu Val Lys Gly Gln Lys
260 265 270

Arg Lys Val Lys Asp Arg Leu Lys Ala Tyr Val Arg Asp Pro Tyr Ala
275 280 285

Leu Asp Leu Ile Asp Lys Leu Leu Val Leu Asp Pro Ala Gln Arg Ile
290 295 300

Asp Ser Asp Asp Ala Leu Asn His Asp Phe Phe Trp Ser Asp Pro Met
305 310 315 320

Pro Ser Asp Leu Lys Gly Met Leu Ser Thr His Leu Thr Ser Met Phe
325 330 335

Glu Tyr Leu Ala Pro Pro Arg Arg Lys Gly Ser Gln Ile Thr Gln Gln
340 345 350

Ser Thr Asn Gln Ser Arg Asn Pro Ala Thr Thr Asn Gln Thr Glu Phe
355 360 365

Glu Arg Val Phe
370

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ACGAATTCCA CACAATCCAA AGATC

25

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CAGAATTCCCT ATTGCCGATC CCCAGA

26

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: one-of(8, 14)
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "N = A or C or G or T"

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 12
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "Y = C or T"

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: one-of(17, 20)
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "R = A or G"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGAATTCNAT GYTNCARCAR CC

22

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: one-of(13, 16, 19, 22, 25)
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "R = A or G"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AACTGCAGTC CARAARAART CRTGRTT

27

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TGTCAAGGAT CAAACCGGCT GTGAT

25

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CGAATTCCAA GAAACGCATC GATGC

25

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AGACCTGCCA AATCGTGT

18

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AGAAGGTGGA TCTGTAACCA TTCGT

25

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GGAATTTCAGA TCTCGATCAG ATTCA

25

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TTACTACTCG AGCTACCAAA CCCGGTC

27

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 28 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TAAGCAAGCT TCTATGGCGC ACATGTCC

28

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TTACTACTCG AGCTACCAAA CCCGGTC

27

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: one-of(13, 16, 22)
- (D) OTHER INFORMATION: /mod_base= OTHER

/note= "Y = C or T"

(ix) FEATURE:

- (A) NAME/KEY: modified_base

(B) LOCATION: 17
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "W = A or T"

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 18
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "S = C or G"

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 19
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "N = A or C or G or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GGAATTCTGG TAYTTYWSNA AYGA

24

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 11
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "Y = C or T"

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 14
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "R = A or G"

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: one-of(17, 20)
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "N = A or C or G or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CGGGATCCTG YTCRAANGGN GGCAT

25

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: one-of(11, 14, 20)
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "N = A or C or G or T"

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 23
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "R = A or G"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CGGGATCCAA NGGNNGCATN CCRT

24

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

ATCACGACAC CACCAGAGCT GTTA

24

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CGAATTCAAGA TCGTGAACGG GA

22

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CGAATTCAAG CGCTAGCAAT G

21

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GAAAGGCGTA GAACCGA

17

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GCTGACCCAT TTCCTGTATC AGATAG

26

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GGAATTCTTC TGCTTGGCGA AT

22

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GGGAATTCGA GGTTCTATAAC ATAT

24

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

CTGTGTGAAT GGAATCTGTG ATGTG

25

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 28 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

TATCCCGGGT CATATGAGTC TCCTAGCC

28

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 13 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Met Leu Gln Gln Pro Ser Gly Ser Thr Pro Ser Asn Val
1 5 10

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 17 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Ala Asp Thr Ala Leu Asn His Asp Phe Phe Trp Thr Asp Pro Met Pro
1 5 10 15

Ser

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Met Leu Gln Gln Pro
1 5

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Asn His Asp Phe Phe Trp Thr
1 5

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Ser Pro Glu Trp Pro Asp Ile
1 5

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Trp Tyr Phe Ser Asn Asp Gln Leu Ala Asn Ser Pro Ser Arg
1 5 10

(2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 14 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Thr Val His Gly Met Pro Pro Phe Glu Gln Gln Leu Pro Tyr
1 5 10

(2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Trp Tyr Phe Ser Asn Asp
1 5

(2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Met Pro Pro Phe Glu Gln
1 5

(2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

His Gly Met Pro Pro Phe
1 5

(2) INFORMATION FOR SEQ ID NO:41:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 41 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GCAGGATCCA GAATTCCATA TGGCAAAGCA GTACGACTCG G

41

(2) INFORMATION FOR SEQ ID NO:42:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

CAGTACTCGA GTTATCAGAA GACGCGCTCA AAC

33

(2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4528 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GGGGGGGGGG GGGTGAATGA AGGAGCGGGC GGAGGGAGGAA TTGTCAATGGC GTCGGGCCGT

60

GGAGCTTCTT CTCGCTGGTT CTCTACTCGG GAACAGCTGG AGAACACGCC GAGCCGCCGC

120

TGCGGAGTGG AGGCGGATAA AGAGCTCTCG TGCCGCCAGC AGGCGGCCAA CCTCATCCAG

180

GAGATGGGAC AGCGTCTCAA TGTCTCTCAG CTTACAATAA ACACTGCGAT TGTTTATATG	240
CACAGGTTTT ATATGCACCA TTCTTCACC AAATTCAACA AAAATATAAT ATCGTCTACT	300
GCATTATTT TGGCTGCAAAGTGGAAAGAA CAGGCTCGAA AACTTGAACA TGTTATCAA	360
GTAGCACATG CTTGTCTTCA TCCTCTAGAG CCACGTGCTGG ATACTAAATG TGATGCTTAC	420
CTTCAACAGA CTCAAGAACT GGTTTACTT GAAACCATAA TGCTACAAAC TCTAGGTTT	480
GAGATCACCA TTGAACACCC ACACACAGAT GTGGTGAAAT GTACCCAGTT AGTAAGAGCA	540
AGCAAGGATT TGGCACAGAC ATCCTATTTC ATGGCTACCA ACAGTCTGCA TCTTACAACC	600
TTCTGTCTTC AGTACAAACC AACAGTGATA GCATGTGTAT GCATTCAATT GGCTTGCAA	660
TGGTCCAATT GGGAGATCCC TGTATCAACT GATGGAAAGC ATTGGTGGGA ATATGTGGAT	720
CCTACAGTTA CTCTAGAATT ATTAGATGAG CTAACACATG AGTTTCTACA AATATTGGAG	780
AAAACGCCTA ATAGGTTGAA GAAGATTGAA AACTGGAGGG CTAATCAGGC AGCTAGGAAA	840
CCAAAAGTAG ATGGACAGGT ATCAGAGACA CCACTTCTTG GTTCATCTTT GGTCCAGAAT	900
TCCATTTAG TAGATAGTGT CACTGGTGTG CCTACAAACC CAAGTTTCA GAAACCATCT	960
ACATCAGCAT TCCCTGCGCC AGTACCTCTA AATTCAAGGAA ATATTCTGT TCAAGACAGC	1020
CATACATCTG ATAATTTGTC AATGCTAGCA ACAGGAATGC CAAGTACTTC ATACGGTTA	1080
TCATCACACC AGGAATGGCC TCAACATCAA GACTCAGCAA GGACAGAACAA GCTATATTCA	1140
CAGAAACAGG AGACATCTT GTCTGGTAGC CAGTACAACA TCAACTTCCA GCAGGGACCT	1200
TCTATATCAC TGCATTCAAGG ATTACATCAC AGACCTGACA AAATTCAAGA TCATTCTTCT	1260
GTAAAGCAAG AATATACTCA TAAAGCAGGG AGCAGTAAAC ACCATGGGCC AATTCCACT	1320
ACTCCAGGAA TAATTCTCA GAAAATGTCT TTAGATAAAAT ATAGAGAAAA GCGTAAACTA	1380
GAAACTCTTG ATCTCGATGT AAGGGATCAT TATATAGCTG CCCAGGTAGA ACAGCAGCAC	1440
AAACAAGGGC AGTCACAGGC AGCCAGCAGC AGTTCTGTTA CTTCTCCCAT TAAAATGAAA	1500
ATACCTATCG CAAATACTGA AAAATACATG GCAGATAAAA AGGAAAAGAG TGGGTCACTG	1560
AAATTACGGA TTCCAATACC ACCCACTGAT AAAAGCGCCA GTAAAGAAGA ACTGAAAATG	1620
AAAATAAAAG TTTCTTCTTC AGAAAGACAC AGCTCTTCTG ATGAAGGCAG TGGGAAAAGC	1680
AAACATTCAA GCCCACATAT TAGCAGAGAC CATAAGGAGA AGCACAAGGA GCATCCTTCA	1740
AGCCGCCACC ACACCAGCAG CCACAAAGCAT TCCCACCTCGC ATAGTGGCAG CAGCAGCGGT	1800

GGCAGTAAAC ACAGTGCCGA CGGAATACCA CCCACTGTT TGAGGAGTCC TGTTGGCCTG 1860
AGCAGTGATG GCATTCCTC TAGCTCCAGC TCTTCAAGGA AGAGGCTGCA TGTCAATGAT 1920
GCATCTCACCA ACCACCACTC CAAAATGAGC AAAAGTTCCA AAAGTCAGG TGGGCTACGG 1980
ACATCTCAGC ACCTCGTGAA ACTGGACAAG AAGCCAGTGG AGACCAACGG TCCTGATGCC 2040
AATCACGAGT ACAGTACAAG CAGCCAGCAT ATGGACTACA AAGACACATT CGACATGCTG 2100
GAECTCACTGT TAAGTGCCCA AGGAATGAAC ATGTAATAAT TTGTTTAGGT CAATTTTCC 2160
TTTACTTTT TAATTTAAAA ATTGTTAGAA TGAAAAAATT CCTTCTGATC TAGCAGTGGT 2220
AACCCCTGCT GTTGCTGCCA CTGCTTCAAT ATTTGTAAGT GCTACTTTAT TCTTCATTCT 2280
GAAAAGAAGA GATTATAGTA AACAAAGTCTT TATCTCCACA TATGATAGTG TTATAAATAC 2340
TGTAAAGGCA TGGAAAGGTGC AAAACTCAGT ATTTCTACAA TTGCAGCTAA GAACATTAGG 2400
ATGAATGGCT GGCTGCTTCT AGGAATATAA GATGCCTCAA GCATTCACTA TTTATGATT 2460
GAATACTGTA GCTATTTTT GTTGCTTGGC TTTTGAATGA GTGTAAATTG TTTTCTTTG 2520
TGTATTTATA CTTGTATGTA TGATTTGCAT GTTTCAATGA TAAAGGGATA AAACAGTATA 2580
CTGACAAC TGTTACAAGAA AGTGGAGAAA ATGTACTACA TTTTGTATGT TTAGATATTA 2640
CCGTAATAC TCAGGATTGG AGCTGCTTGT AAGTATAACA ATATAACAGAA TACTTTATTT 2700
TATCTTGTCA GAGTTCCATC ACTATCTAAA ACAAAAGGTGC AATTTTTAT GTTAACCTTA 2760
AATCTAGCCC TTACTGGAAG CCACTGATAG GGACATTAC TACCAGATGT GTGCAGTGCA 2820
GCAGATGGTC ATATAACACT GTGAGGCACT GAATTTGCC TTCAGAGGTT CTGACCAGAT 2880
TGGCTGCTGA AATAGCCCCT AACTTCTGA AGGCTTGAAG AGGAAAAAAT AAAGTTACA 2940
TACTCTTGAT GGAAGTGCAT TTAAATGTT GTTGGCTTGT TGCAGTTCTA TGAAACAGAG 3000
CTGTTAATAA TGGTTATGTG GATTACTGTG ATTTGAAAAC TAAATTACCA ATAACCTTACC 3060
TAGTAGAGAT TTAGTGAGTT GTTCCCTTTA AAGAATTTTA CACTACATAT TTTAATAGTA 3120
AACAGGGTCA CTTTCCTTTA GCATTCAAGAA TGACACCATA TTCTTAAATA TACTCCTTCC 3180
CTGAAGCGTG TTTGTGTGTG ATGCCATATT TCTTTTCAG GTAAATGTAG TCTTCCTTAT 3240
AAAAATGAAA TTAAACCTAT GCTCTCAATT CTTTTATATT CTAACAATAA ATAAAAAAGA 3300
AAAGATTACT GACTGTGCAT TGTACCTGTA TTTATAGTTT ATGGTTATCA GAAGCTCTGT 3360
AAGAAAGAAA AGGTCAGCTC CCAGGCAAAC CAGTAGTGGA GGTTTACAT TTGTTGCAC 3420

DRAFT

ATCTCAGTAT ATTTCTGTTG AGGTAAAGTT TGCACAGTCA TCTGACTTCT GATCAAGCAT	3480
TAGATTTAA CTTGTTAGA TTTGTCCTTA AACACCAGTA ATATGGCTCT TGTTTATCAG	3540
CTAATCTTGA ATTTATTCTG TGGTAAATCT TTTGAGTTGC TGAGTATATT TGAGATTGAT	3600
TGGATTCAAC CTCTTGTGA ACTGAAAACT TAATTTTTC TCTGTATTT TGTTACAAAG	3660
CCACTGATAC GTGCACAATT GTAATTAAGT ATGTTGCAGT TGTAAATATT AGAGTTAAC	3720
CTCATGCTCT ACCTTTATTT AGCAATTACC TAATTTGCCA GTAGCTTAT AATTTTAA	3780
GATAATTGTT CATTATTTG TCAATGTTAT TTGAACCTGG GGTACTTAGG AGCCTCTTG	3840
TAGGGACTGT GCCTAGGTAG CATGTCTAA CATTGTTCT GGTCTTGCAT AACCTCAGTA	3900
TCTTGTCAT TATATGTAAC TTTGTTGCTC TGTATGGCAT AATATTGTAT CCATAAACAT	3960
GGTAATTGTT ATACAGTTAT ACTTTACAG TGGTACATAA TCCAAGGACT AGTATAGAAT	4020
TAAGCTGAGT GCAAGATGAG GGAGGGAAAGG GCTTCTTGG TAATTTAGAT GTGAAACCTC	4080
TACAGAGCTA TCATGTAAAA ACTACATGAG GTGGTTGTGC TACTGTATAA TTGGGGGTGA	4140
TAATACCAGG AATTTAATA AGATTTGTA AAGAATATCC AGAAAAGTAG TGAACCTTATT	4200
TTCAGTAGGC ATAGAAAACA ATGTGAATAT TTAAGGTCTG TGACTATAGT TAAACTTCAC	4260
TAAGAATTG CAGAATTGTT TTGAGATGTG TGAATAAAGG TAATTTATT GAATCTTCAT	4320
TGGTGCTAAT GTTGGACAGT TAAAAAGATA GCTAGTGTAT ATTGTTATGG GTCAGTACTT	4380
ATTAGTACTT CCAAAATTGA ATTTGAAATG CTATGTATTC ACTTTCACT CTGTAAATGT	4440
AATTCTTAC AATGACTTTA TTTATTAAG GGCAGCCAGT TGTCAATTGT AAAAAAAAAA	4500
AAAAAAAAAA AAAGCGGCCG CTGAATTC	4528

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2091 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

ATGGCGTCGG GCCGTGGAGC TTCTTCTCGC TGTTCTTTA CTCGGAAACA GCTGGAGAAC	60
ACGCCGAGCC GCCGCTGCGG AGTGGAGGCG GATAAAGAGC TCTCGTGCCTG CCAGCAGGCG	120

PGC914896_257029

GCCAACCTCA TCCAGGAGAT GGGACAGCGT CTCAGCTTAC AATAAACACT	180
GCGATTGTTT ATATGCACAG GTTTTATATG CACCATTCTT TCACCAAATT CAACAAAAAT	240
ATAATATCGT CTACTGCATT ATTTTGCT GCAGAAAGTGG AAGAACAGGC TCGAAAAC	300
GAACATGTTA TCAAAGTAGC ACATGCTTGT CTTCATCCTC TAGAGCCACT GCTGGATACT	360
AAATGTGATG CTTACCTTCA ACAGACTCAA GAACGGTTA TACTTGAAAC CATAATGCTA	420
CAAACCTCTAG GTTTGAGAT CACCATTGAA CACCCACACA CAGATGTGGT GAAATGTACC	480
CAGTTAGTAA GAGCAAGCAA GGATTTGGCA CAGACATCCT ATTTCATGGC TACCAACAGT	540
CTGCATCTTA CAACCTTCTG TCTTCAGTAC AAACCAACAG TGATAGCATG TGTATGCATT	600
CATTTGGCTT GCAAATGGTC CAATTGGAG ATCCCTGTAT CAACTGATGG AAAGCATTGG	660
TGGGAATATG TGGATCCTAC AGTTACTCTA GAATTATTAG ATGAGCTAAC ACATGAGTTT	720
CTACAAATAT TGGAGAAAAC GCCTAATAGG TTGAAGAAGA TTCGAAACTG GAGGGCTAAT	780
CAGGCAGCTA GGAAACCAAA AGTAGATGGA CAGGTATCAG AGACACCACT TCTTGGTTCA	840
TCTTTGGTCC AGAATTCCAT TTTAGTAGAT AGTGTCACTG GTGTGCCTAC AAACCCAAGT	900
TTTCAGAAC CATCTACATC AGCATTCCCT GCGCCAGTAC CTCTAAATTC AGGAAATATT	960
TCTGTTCAAG ACAGCCATAC ATCTGATAAT TTGTCATGC TAGCACAGG AATGCCAAGT	1020
ACTTCATACG GTTTATCATC ACACCAGGAA TGGCCTCAAC ATCAAGACTC AGCAAGGACA	1080
GAACAGCTAT ATTACAGAA ACAGGAGACA TCTTTGTCTG GTAGCCAGTA CAACATCAAC	1140
TTCCAGCAGG GACCTTCTAT ATCACTGCAT TCAGGATTAC ATCACAGACC TGACAAAATT	1200
TCAGATCATT CTTCTGTTAA GCAAGAATAT ACTCATAAAG CAGGGAGCAG TAAACACCAT	1260
GGGCCAATT CCACTAATCC AGGAATAATT CCTCAGAAAA TGTCTTCTA TAAATATAGA	1320
GAAAAGCGTA AACTAGAAC TCTTGATCTC GATGTAAGGG ATCATTATAT AGCTGCCAG	1380
GTAGAACAGC AGCACAAACA AGGGCAGTCA CAGGCAGCCA GCAGCAGTTC TGTTACTTCT	1440
CCCATTAAAA TGAAAATACC TATCGCAAAT ACTGAAAAT ACATGGCAGA TAAAAGGAA	1500
AAGAGTGGGT CACTGAAATT ACGGATTCCA ATACCACCCA CTGATAAAAG CGCCAGTAA	1560
GAAGAACTGA AAATGAAAAT AAAAGTTCT TCTTCAGAAA GACACAGCTC TTCTGATGAA	1620
GGCAGTGGGA AAAGCAAACA TTCAAGCCCA CATATTAGCA GAGACCATAA GGAGAACGAC	1680
AAGGAGCAGTC CTTCAAGCCG CCACCACACC AGCAGCCACA AGCATTCCCA CTCGCATAGT	1740

GGCAGCAGCA GCGGTGGCAG TAAACACAGT GCCGACGGAA TACCACCCAC	TGTTCTGAGG	1800
AGTCCTGTTG GCCTGAGCAG TGATGGCATT TCCTCTAGCT CCAGCTCTTC	AAGGAAGAGG	1860
CTGCATGTCA ATGATGCATC TCACAACCAC CACTCCAAAA TGAGCAAAAG	TTCCAAAAGT	1920
TCAGGTGGGC TACGGACATC TCAGCACCTC GTGAAACTGG ACAAGAAGCC	AGTGGAGACC	1980
AACGGTCCTG ATGCCAATCA CGAGTACAGT ACAAGCAGCC AGCATATGGA	CTACAAAGAC	2040
ACATTCGACA TGCTGGACTC ACTGTTAAGT GCCCAAGGAA TGAACATGTA A		2091

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 696 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

Met Ala Ser Gly Arg Gly Ala Ser Ser Arg Trp Phe Phe Thr Arg Glu			
1	5	10	15
Gln Leu Glu Asn Thr Pro Ser Arg Arg Cys Gly Val Glu Ala Asp Lys			
20	25	30	
Glu Leu Ser Cys Arg Gln Gln Ala Ala Asn Leu Ile Gln Glu Met Gly			
35	40	45	
Gln Arg Leu Asn Val Ser Gln Leu Thr Ile Asn Thr Ala Ile Val Tyr			
50	55	60	
Met His Arg Phe Tyr Met His His Ser Phe Thr Lys Phe Asn Lys Asn			
65	70	75	80
Ile Ile Ser Ser Thr Ala Leu Phe Leu Ala Ala Lys Val Glu Glu Gln			
85	90	95	
Ala Arg Lys Leu Glu His Val Ile Lys Val Ala His Ala Cys Leu His			
100	105	110	
Pro Leu Glu Pro Leu Leu Asp Thr Lys Cys Asp Ala Tyr Leu Gln Gln			
115	120	125	
Thr Gln Glu Leu Val Ile Leu Glu Thr Ile Met Leu Gln Thr Leu Gly			
130	135	140	
Phe Glu Ile Thr Ile Glu His Pro His Thr Asp Val Val Lys Cys Thr			
145	150	155	160
Gln Leu Val Arg Ala Ser Lys Asp Leu Ala Gln Thr Ser Tyr Phe Met			

165 170 175

Ala Thr Asn Ser Leu His Leu Thr Thr Phe Cys Leu Gln Tyr Lys Pro
180 185 190

Thr Val Ile Ala Cys Val Cys Ile His Leu Ala Cys Lys Trp Ser Asn
195 200 205

Trp Glu Ile Pro Val Ser Thr Asp Gly Lys His Trp Trp Glu Tyr Val
210 215 220

Asp Pro Thr Val Thr Leu Glu Leu Leu Asp Glu Leu Thr His Glu Phe
225 230 235 240

Leu Gln Ile Leu Glu Lys Thr Pro Asn Arg Leu Lys Lys Ile Arg Asn
245 250 255

Trp Arg Ala Asn Gln Ala Ala Arg Lys Pro Lys Val Asp Gly Gln Val
260 265 270

Ser Glu Thr Pro Leu Leu Gly Ser Ser Leu Val Gln Asn Ser Ile Leu
275 280 285

Val Asp Ser Val Thr Gly Val Pro Thr Asn Pro Ser Phe Gln Lys Pro
290 295 300

Ser Thr Ser Ala Phe Pro Ala Pro Val Pro Leu Asn Ser Gly Asn Ile
305 310 315 320

Ser Val Gln Asp Ser His Thr Ser Asp Asn Leu Ser Met Leu Ala Thr
325 330 335

Gly Met Pro Ser Thr Ser Tyr Gly Leu Ser Ser His Gln Glu Trp Pro
340 345 350

Gln His Gln Asp Ser Ala Arg Thr Glu Gln Leu Tyr Ser Gln Lys Gln
355 360 365

Glu Thr Ser Leu Ser Gly Ser Gln Tyr Asn Ile Asn Phe Gln Gln Gly
370 375 380

Pro Ser Ile Ser Leu His Ser Gly Leu His His Arg Pro Asp Lys Ile
385 390 395 400

Ser Asp His Ser Ser Val Lys Gln Glu Tyr Thr His Lys Ala Gly Ser
405 410 415

Ser Lys His His Gly Pro Ile Ser Thr Thr Pro Gly Ile Ile Pro Gln
420 425 430

Lys Met Ser Leu Asp Lys Tyr Arg Glu Lys Arg Lys Leu Glu Thr Leu
435 440 445

Asp Leu Asp Val Arg Asp His Tyr Ile Ala Ala Gln Val Glu Gln Gln

450 455 460

His	Lys	Gln	Gly	Gln	Ser	Gln	Ala	Ala	Ser	Ser	Ser	Val	Thr	Ser	
465														480	
Pro	Ile	Lys	Met	Lys	Ile	Pro	Ile	Ala	Asn	Thr	Glu	Lys	Tyr	Met	Ala
														495	
Asp	Lys	Lys	Glu	Lys	Ser	Gly	Ser	Leu	Lys	Leu	Arg	Ile	Pro	Ile	Pro
														510	
Pro	Thr	Asp	Lys	Ser	Ala	Ser	Lys	Glu	Glu	Leu	Lys	Met	Lys	Ile	Lys
														515	
515								520						525	
Val	Ser	Ser	Ser	Glu	Arg	His	Ser	Ser	Ser	Asp	Glu	Gly	Ser	Gly	Lys
														530	
530								535						540	
Ser	Lys	His	Ser	Ser	Pro	His	Ile	Ser	Arg	Asp	His	Lys	Glu	Lys	His
														545	
545								550						555	
Lys	Glu	His	Pro	Ser	Ser	Arg	His	His	Thr	Ser	Ser	His	Lys	His	Ser
														560	
565									570					575	
His	Ser	His	Ser	Gly	Ser	Ser	Gly	Gly	Ser	Lys	His	Ser	Ala	Asp	
														580	
580								585						590	
Gly	Ile	Pro	Pro	Thr	Val	Leu	Arg	Ser	Pro	Val	Gly	Leu	Ser	Ser	Asp
														595	
595								600						605	
Gly	Ile	Ser	Ser	Ser	Ser	Ser	Ser	Arg	Lys	Arg	Leu	His	Val	Asn	
														610	
610								615						620	
Asp	Ala	Ser	His	Asn	His	His	Ser	Lys	Met	Ser	Lys	Ser	Ser	Lys	Ser
														625	
625								630						640	
Ser	Gly	Gly	Leu	Arg	Thr	Ser	Gln	His	Leu	Val	Lys	Leu	Asp	Lys	Lys
														645	
645									650					655	
Pro	Val	Glu	Thr	Asn	Gly	Pro	Asp	Ala	Asn	His	Glu	Tyr	Ser	Thr	Ser
														660	
660									665					670	
Ser	Gln	His	Met	Asp	Tyr	Lys	Asp	Thr	Phe	Asp	Met	Leu	Asp	Ser	Leu
														675	
675								680						685	
Leu	Ser	Ala	Gln	Gly	Met	Asn	Met								
														690	
690								695							

(2) INFORMATION FOR SEQ ID NO:46:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2190 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

ATGGCGTCGG	GCCGTGGAGC	TTCTTCTCGC	TGTTCTTA	CTCGGAACA	GCTGGAGAAC	60
ACGCCGAGCC	GCCGCTCGG	AGTGGAGGCG	GATAAAGAGC	TCTCGTGC	CCAGCAGGCG	120
GCCAACCTCA	TCCAGGAGAT	GGGACAGCGT	CTCAATGTCT	CTCAGCTTAC	AATAAACACT	180
GCGATTGTTT	ATATGCACAG	GTTTATATG	CACCATTCTT	TCACCAAATT	CAACAAAAAT	240
ATAATATCGT	CTACTGCATT	ATTTTGCT	GCAAAAGTGG	AAGAACAGGC	TCGAAAAC	300
GAACATGTTA	TCAAAGTAGC	ACATGCTTGT	CTTCATCCTC	TAGAGCCACT	GCTGGATACT	360
AAATGTGATG	CTTACCTTCA	ACAGACTCAA	GAACTGGTTA	TACTTGAAAC	CATAATGCTA	420
CAAACCTCTAG	GTTTGAGAT	CACCATTGAA	CACCCACACA	CAGATGTGGT	GAAATGTACC	480
CAGTTAGTAA	GAGCAAGCAA	GGATTTGGCA	CAGACATCCT	ATTTCATGGC	TACCAACAGT	540
CTGCATCTTA	CAACCTTCTG	TCTTCAGTAC	AAACCAACAG	TGATAGCATG	TGTATGCATT	600
CATTTGGCTT	GCAAATGGTG	CAATTGGAG	ATCCCTGTAT	CAACTGATGG	AAAGCATTGG	660
TGGGAATATG	TGGATCCTAC	AGTTACTCTA	GAATTATTAG	ATGAGCTAAC	ACATGAGTTT	720
CTACAAATAT	TGGAGAAAAC	GCCTAATAGG	TTGAAGAAGA	TTCGAAACTG	GAGGGCTAAT	780
CAGGCAGCTA	GGAAACCAAA	AGTAGATGGA	CAGGTATCAG	AGACACCACT	TCTTGGTTCA	840
TCTTTGGTCC	AGAATTCCAT	TTTAGTAGAT	AGTGTCACTG	GTGTGCCTAC	AAACCCAAGT	900
TTTCAGAAAC	CATCTACATC	AGCATTCCCT	GCGCCAGTAC	CTCTAAATTC	AGGAAATATT	960
TCTGTTCAAG	ACAGCCATAC	ATCTGATAAT	TTGTCATGC	TAGCAACAGG	AATGCCAAGT	1020
ACTTCATACG	GTTCATCATC	ACACCAGGA	TGGCCTCAAC	ATCAAGACTC	AGCAAGGACA	1080
GAACAGCTAT	ATTCACAGAA	ACAGGAGACA	TCTTGTCTG	GTAGCCAGTA	CAACATCAAC	1140
TTCCAGCAGG	GACCTTCTAT	ATCACTGCAT	TCAGGATTAC	ATCACAGACC	TGACAAAATT	1200
TCAGATCATT	CTTCTGTTAA	GCAGGAATAT	ACTCATAAAG	CAGGGAGCAG	TAAACACCAT	1260
GGGCCAATT	CCACTACTCC	AGGAATAATT	CCTCAGAAAA	TGTCTTACA	TAAATATAGA	1320
GAAAAGCGTA	AACTAGAAAC	TCTTGATCTC	GATGTAAGGG	ATCATTATAT	AGCTGCCAG	1380
GTAGAACAGC	AGCACAAACA	AGGGCAGTCA	CAGGCAGCCA	GCAGCAGTTC	TGTTACTTCT	1440
CCCATTAAAA	TGAAAATACC	TATCGCAAAT	ACTGAAAAAT	ACATGGCAGA	AAAAAGGAA	1500

AAGAGTGGGT CACTGAAATT ACGGATTCCA ATACCACCCA CTGATAAAAG CGCCAGTAAA	1560
GAAGAACTGA AAATGAAAAT AAAAGTTCT TCTTCAGAAA GACACAGCTC TTCTGATGAA	1620
GGCAGTGGGA AAAGCAAACA TTCAAGCCC CATATTAGCA GAGACCATAA GGAGAACGCAC	1680
AAGGAGCATC CTTCAAGCCG CCACCACACC AGCAGCCACA AGCATTCCC CTCGCATAGT	1740
GGCAGCAGCA GCGGTGGCAG TAAACACAGT GCCGACGGAA TACCACCCAC TGTTCTGAGG	1800
AGTCCTGTTG GCCTGAGCAG TGATGGCATT TCCTCTAGCT CCAGCTCTTC AAGGAAGAGG	1860
CTGCATGTCA ATGATGCATC TCACAACCAC CACTCCAAA TGAGCAAAAG TTCCAAAAGT	1920
TCAGGTAGTT CATCTAGTTC TTCCTCCTCT GTTAAGCAGT ATATATCCTC TCACAACTCT	1980
GTTTTAACCC ATCCCTTACC CCTCCTCCCC TGTCACATAC CAGGTGGGCT ACGGACATCT	2040
CTGCACCTCG TGAAACTGGA CAAGAAGCCA GTGGAGACCA ACGGTCCCTGA TGCCAATCAC	2100
GAGTACAGTA CAAGCAGCCA GCATATGGAC TACAAAGACA CATTGACAT GCTGGACTCA	2160
CTGTTAAGTG CCCAAGGAAT GAACATGTAA	2190

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 729 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

Met Ala Ser Gly Arg Gly Ala Ser Ser Arg Trp Phe Phe Thr Arg Glu			
1	5	10	15

Gln Leu Glu Asn Thr Pro Ser Arg Arg Cys Gly Val Glu Ala Asp Lys			
20	25	30	

Glu Leu Ser Cys Arg Gln Gln Ala Ala Asn Leu Ile Gln Glu Met Gly			
35	40	45	

Gln Arg Leu Asn Val Ser Gln Leu Thr Ile Asn Thr Ala Ile Val Tyr			
50	55	60	

Met His Arg Phe Tyr Met His His Ser Phe Thr Lys Phe Asn Lys Asn			
65	70	75	80

Ile Ile Ser Ser Thr Ala Leu Phe Leu Ala Ala Lys Val Glu Glu Gln			
85	90	95	

Ala Arg Lys Leu Glu His Val Ile Lys Val Ala His Ala Cys Leu His	
---	--

100 105 110

Pro Leu Glu Pro Leu Leu Asp Thr Lys Cys Asp Ala Tyr Leu Gln Gln
115 120 125

Thr Gln Glu Leu Val Ile Leu Glu Thr Ile Met Leu Gln Thr Leu Gly
130 135 140

Phe Glu Ile Thr Ile Glu His Pro His Thr Asp Val Val Lys Cys Thr
145 150 155 160

Gln Leu Val Arg Ala Ser Lys Asp Leu Ala Gln Thr Ser Tyr Phe Met
165 170 175

Ala Thr Asn Ser Leu His Leu Thr Thr Phe Cys Leu Gln Tyr Lys Pro
180 185 190

Thr Val Ile Ala Cys Val Cys Ile His Leu Ala Cys Lys Trp Ser Asn
195 200 205

Trp Glu Ile Pro Val Ser Thr Asp Gly Lys His Trp Trp Glu Tyr Val
210 215 220

Asp Pro Thr Val Thr Leu Glu Leu Leu Asp Glu Leu Thr His Glu Phe
225 230 235 240

Leu Gln Ile Leu Glu Lys Thr Pro Asn Arg Leu Lys Lys Ile Arg Asn
245 250 255

Trp Arg Ala Asn Gln Ala Ala Arg Lys Pro Lys Val Asp Gly Gln Val
260 265 270

Ser Glu Thr Pro Leu Leu Gly Ser Ser Leu Val Gln Asn Ser Ile Leu
275 280 285

Val Asp Ser Val Thr Gly Val Pro Thr Asn Pro Ser Phe Gln Lys Pro
290 295 300

Ser Thr Ser Ala Phe Pro Ala Pro Val Pro Leu Asn Ser Gly Asn Ile
305 310 315 320

Ser Val Gln Asp Ser His Thr Ser Asp Asn Leu Ser Met Leu Ala Thr
325 330 335

Gly Met Pro Ser Thr Ser Tyr Gly Leu Ser Ser His Gln Glu Trp Pro
340 345 350

Gln His Gln Asp Ser Ala Arg Thr Glu Gln Leu Tyr Ser Gln Lys Gln
355 360 365

Glu Thr Ser Leu Ser Gly Ser Gln Tyr Asn Ile Asn Phe Gln Gln Gly
370 375 380

Pro Ser Ile Ser Leu His Ser Gly Leu His His Arg Pro Asp Lys Ile

385 390 395 400

Ser Asp His Ser Ser Val Lys Gln Glu Tyr Thr His Lys Ala Gly Ser
405 410 415

Ser Lys His His Gly Pro Ile Ser Thr Thr Pro Gly Ile Ile Pro Gln
420 425 430

Lys Met Ser Leu Asp Lys Tyr Arg Glu Lys Arg Lys Leu Glu Thr Leu
435 440 445

Asp Leu Asp Val Arg Asp His Tyr Ile Ala Ala Gln Val Glu Gln Gln
450 455 460

His Lys Gln Gly Gln Ser Gln Ala Ala Ser Ser Ser Ser Val Thr Ser
465 470 475 480

Pro Ile Lys Met Lys Ile Pro Ile Ala Asn Thr Glu Lys Tyr Met Ala
485 490 495

Asp Lys Lys Glu Lys Ser Gly Ser Leu Lys Leu Arg Ile Pro Ile Pro
500 505 510

Pro Thr Asp Lys Ser Ala Ser Lys Glu Glu Leu Lys Met Lys Ile Lys
515 520 525

Val Ser Ser Ser Glu Arg His Ser Ser Ser Asp Glu Gly Ser Gly Lys
530 535 540

Ser Lys His Ser Ser Pro His Ile Ser Arg Asp His Lys Glu Lys His
545 550 555 560

Lys Glu His Pro Ser Ser Arg His His Thr Ser Ser His Lys His Ser
565 570 575

His Ser His Ser Gly Ser Ser Ser Gly Gly Ser Lys His Ser Ala Asp
580 585 590

Gly Ile Pro Pro Thr Val Leu Arg Ser Pro Val Gly Leu Ser Ser Asp
595 600 605

Gly Ile Ser Ser Ser Ser Ser Ser Arg Lys Arg Leu His Val Asn
610 615 620

Asp Ala Ser His Asn His His Ser Lys Met Ser Lys Ser Ser Lys Ser
625 630 635 640

Ser Gly Ser Ser Ser Ser Ser Ser Val Lys Gln Tyr Ile Ser
645 650 655

Ser His Asn Ser Val Phe Asn His Pro Leu Pro Leu Leu Pro Cys His
660 665 670

Ile Pro Gly Gly Leu Arg Thr Ser Gln His Leu Val Lys Leu Asp Lys

675	680	685
Lys Pro Val Glu Thr Asn Gly Pro Asp Ala Asn His Glu Tyr Ser Thr		
690	695	700
Ser Ser Gln His Met Asp Tyr Lys Asp Thr Phe Asp Met Leu Asp Ser		
705	710	715
Leu Leu Ser Ala Gln Gly Met Asn Met		
725		

(2) INFORMATION FOR SEQ ID NO:48:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2360 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GGAAGTGCCT GCAACCTTCG CCGCTGCCCTT CTGGTTGAAG CACTATGGAG GGAGAGAGGA	60
AGAACAAACAA CAAACGGTGG TATTCACTC GAGAACAGCT GGAAAATAGC CCATCCCGTC	120
GTTTGCGGT GGACCCAGAT AAAGAACATTCTT CTTATGCCA GCAGGGGCC AATCTGCTTC	180
AGGACATGGG GCAGCGTCTT AACGTCTCAC AATTGACTAT CAACACTGCT ATAGTATAACA	240
TGCATCGATT CTACATGATT CAGTCCTTCA CACGGTTCCC TGAAATTCT GTGGCTCCAG	300
CAGCCTTGTT TCTAGCAGCT AAAGTGGAGG AGCAGCCCAA AAAATTGGAA CATGTCATCA	360
AGGTAGCACA TACTTGTCTC CATCCTCAGG AATCCCTTCC TGATACTAGA AGTGAGGCTT	420
ATTTGCAACA AGTTCAAGAT CTGGTCATTT TAGAAAGCAT AATTTGCAG ACTTTAGGCT	480
TTGAACTAAC AATTGATCAC CCACATACTC ATGTAGTAAA GTGCACTCAA CTTGTTCGAG	540
CAAGCAAGGA CTTAGCACAG ACTTCTTAAC TCATGGCAAC CAACAGCCTG CATTGACCA	600
CATTTAGCCT GCAGTACACA CCTCCTGTGG TGGCCTGTGT CTGCATTACAC CTGGCTTGCA	660
AGTGGTCCAA TTGGGAGATC CCAGTCTCAA CTGACGGGAA GCACTGGTGG GAGTATGTTG	720
ACGCCACTGT GACCTTGGAA CTTTAGATG AACTGACACA TGAGTTCTA CAGATTTGG	780
AGAAAACCTCC CAACAGGCTC AAACGCATTT GGAATTGGAG GGCATGCGAG GCTGCCAAGA	840
AAACAAAAGC AGATGACCGA GGAACAGATG AAAAGACTTC AGAGCAGACA ATCCTCAATA	900
TGATTTCCA GAGCTCTTCA GACACAACCA TTGCAGGTTT AATGAGCATG TCAACTTCTA	960

PR095172
2023-07-10 10:23:45

CCACAAGTGC	AGTGCCTTCC	CTGCCAGTCT	CCGAAGAGTC	ATCCAGCAAC	TTAACCGATG	1020
TGGAGATGTT	GCCGGGCAAG	CGTTGGCTGT	CCTCCAACC	TTCTTCAAA	CTAGAACCTA	1080
CTCAGGGTCA	TCGGACTAGT	GAGAATTAG	CACTTACAGG	AGTTGATCAT	TCCTTACAC	1140
AGGATGGTTC	AAATGCATTT	ATTTCCCAGA	AGCAGAATAG	TAAGAGTGTG	CCATCAGCTA	1200
AAGTGTCACT	GAAAGAATAC	CGCGCGAAGC	ATGCAGAAGA	ATTGGCTGCC	CAGAAGAGGC	1260
AACTGGAGAA	CATGGAAGCC	AATGTGAAGT	CACAATATGC	ATATGCTGCC	CAGAATCTCC	1320
TTTCTCATCA	TGATAGCCAT	TCTTCAGTCA	TTCTAAAAAT	GCCCCATAGAG	GGTCAGAAA	1380
ACCCCGAGCG	GCCTTTCTG	GAAAAGGCTG	ACAAAACAGC	TCTCAAAATG	AGAATCCCAG	1440
TGGCAGGTGG	AGATAAAAGCT	GCGTCTCAA	AACCAGAGGA	GATAAAAATG	CCGCATAAAAG	1500
TCCATGCTGC	AGCTGATAAG	CACAATTCTG	TAGAGGACAG	TGTTACAAAG	AGCCGAGAGC	1560
ACAAAGAAGA	GCGCAAGACT	CACCCATCTA	ATCATCATCA	TCATCATAAT	CACCACTCAC	1620
ACAAGCACTC	TCATTCCAA	CTTCCAGTTG	GTACTGGAA	CAAACGTCCT	GGTGATCCAA	1680
AACATAGTAG	CCAGACAAGC	AACTTAGCAC	ATAAAACCTA	TAGCTGTCT	AGTTCTTTT	1740
CCTCTTCCAG	TTCTACTCGT	AAAAGGGGAC	CCTCTGAAGA	GACTGGAGGG	GCTGTGTTG	1800
ATCATCCAGC	CAAGATTGCC	AAGAGTACTA	AATCCTCTTC	CCTAAATTTC	TCCTTCCCTT	1860
CACTTCCTAC	AATGGGTCAG	ATGCCTGGC	ATAGCTCAGA	CACAAGTGGC	CTTTCCCTTT	1920
CACAGCCCAG	CTGTAAAACT	CGTGTCCCTC	ATTCGAAACT	GGATAAAGGG	CCCACTGGGG	1980
CCAATGGTCA	CAACACGACC	CAGACAATAG	ACTATCAAGA	CACTGTGAAT	ATGCTTCACT	2040
CCCTGCTCAG	TGCCCAGGGT	GTTCAGCCC	CTCAGCCCAC	TGCATTGAA	TTTGTTCGTC	2100
CTTATAGTGA	CTATCTGAAT	CCTCGGTCTG	GTGGAATCTC	CTCGAGATCT	GGCAATACAG	2160
ACAAACCCCG	GCCACCACCT	CTGCCATCAG	AACCTCCTCC	ACCACCTCCA	CCCCTTCCTA	2220
AGTAAAAAAA	GAAAAAGAAG	AGGAGAAAAA	AACTTCTTA	AAAAAACACA	TAATTTTCT	2280
TTTTTTTTTG	GGGAAAAAAA	AATTTTTTT	AAAATTTTT	CCCCAAGGGA	CGGGGGAAAA	2340
TTTTATTTTT	AAAATTTTT					2360

(2) INFORMATION FOR SEQ ID NO:49:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2181 base pairs
 - (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

ATGGAGGGAG AGAGGAAGAA CAACAACAAA CGGTGGTATT TCACTCGAGA ACAGCTGGAA	60
AATAGCCCAT CCCGTCGTTT TGGCGTGGAC CCAGATAAAG AACTTTCTTA TCGCCAGCAG	120
GCGGCCAAC TGCTTCAGGA CATGGGCAG CGTCTTAACG TCTCACAAATT GACTATCAAC	180
ACTGCTATAG TATAATGCA TCGATTCTAC ATGATTCACT CCTTCACACG GTTCCCTGGA	240
AATTCTGTGG CTCCAGCAGC CTTGTTCTA GCAGCTAAAG TGGAGGAGCA GCCCAAAAAA	300
TTGGAACATG TCATCAAGGT AGCACATACT TGTCTCCATC CTCAGGAATC CCTTCCTGAT	360
ACTAGAAGTG AGGCTTATTT GCAACAAGTT CAAGATCTGG TCATTTAGA AAGCATAATT	420
TTGCAGACTT TAGGCTTGA ACTAACAAATT GATCACCCAC ATACTCATGT AGTAAAGTGC	480
ACTCAACTTG TTCGAGCAAG CAAGGACTTA GCACAGACTT CTTACTTCAT GGCAACCAAC	540
AGCCTGCATT TGACCACATT TAGCCTGCAG TACACACCTC CTGTGGTGGC CTGTGTCTGC	600
ATTCACCTGG CTTGCAAGTG GTCCAATTGG GAGATCCCAG TCTCACTGA CGGGAAGCAC	660
TGGTGGAGT ATGTTGACGC CACTGTGACC TTGGAACCTT TAGATGAAC GACACATGAG	720
TTTCTACAGA TTTGGAGAA AACTCCAAC AGGCTAAAC GCATTGGAA TTGGAGGGCA	780
TGCGAGGCTG CCAAGAAAAC AAAAGCAGAT GACCGAGGAA CAGATGAAAA GACTTCAGAG	840
CAGACAATCC TCAATATGAT TTCCCAGAGC TCTTCAGACA CAACCATTGC AGGTTTAATG	900
AGCATGTCAA CTTCTACCAC AAGTGCAGTG CCTTCCCTGC CAGTCTCCGA AGAGTCATCC	960
AGCAACTTAA CCAGTGTGGA GATGTTGCCG GGCAAGCGTT GGCTGTCCTC CCAACCTTCT	1020
TTCAAACTAG AACCTACTCA GGGTCATCGG ACTAGTGAGA ATTTAGCACT TACAGGAGTT	1080
GATCATTCCCT TACCACAGGA TGGTTCAAAT GCATTTATTT CCCAGAAGCA GAATAGTAAG	1140
AGTGTGCCAT CAGCTAAAGT GTCACTGAAA GAATACCGCG CGAACATGC AGAAGAATTG	1200
GCTGCCAGA AGAGGCAACT GGAGAACATG GAAGCCAATG TGAAGTCACA ATATGCATAT	1260
GCTGCCAGA ATCTCCTTTC TCATCATGAT AGCCATTCTT CAGTCATTCT AAAAATGCC	1320
ATAGAGGGTT CAGAAAACCC CGAGCGGCCT TTTCTGGAAA AGGCTGACAA AACAGCTCTC	1380
AAAATGAGAA TCCCAGTGGC AGGTGGAGAT AAAGCTGCGT CTTCAAAACC AGAGGAGATA	1440
AAAATGCGCA TAAAAGTCCA TGCTGCAGCT GATAAGCACA ATTCTGTAGA GGACAGTGT	1500

ACAAAAGAGCC GAGAGCACAA AGAAGAGCGC AAGACTCACC CATCTAATCA TCATCATCAT 1560
 CATAATCACC ACTCACACAA GCACTCTCAT TCCCAACTTC CAGTTGGTAC TGGGAACAAA 1620
 CGTCCTGGTG ATCCAAAACA TAGTAGCCAG ACAAGCAACT TAGCACATAA AACCTATAGC 1680
 TTGTCTAGTT CTTTTTCCTC TTCCAGTTCT ACTCGTAAAA GGGGACCCTC TGAAGAGACT 1740
 GGAGGGGCTG TGTTTGATCA TCCAGCCAAG ATTGCCAAGA GTACTAAATC CTCTTCCCTA 1800
 AATTTCCTCCT TCCCTTCACT TCCTACAATG GGTCAAGATGC CTGGGCATAG CTCAGACACA 1860
 AGTGGCCTTT CCTTTTCACA GCCCAGCTGT AAAACTCGTG TCCCTCATTC GAAACTGGAT 1920
 AAAGGGCCCA CTGGGGCCAA TGGTCACAAC ACGACCCAGA CAATAGACTA TCAAGACACT 1980
 GTGAATATGC TTCACTCCCT GCTCAGTGCC CAGGGTGTTC AGCCCCTCA GCCCACTGCA 2040
 TTTGAATTTG TTCGTCCCTTA TAGTGACTAT CTGAATCCTC GGTCTGGTGG AATCTCCTCG 2100
 AGATCTGGCA ATACAGACAA ACCCCGGCCA CCACCTCTGC CATCAGAACC TCCTCCACCA 2160
 CTTCCACCCCC TTCCTAAGTA A 2181

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 726 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

Met	Glu	Gly	Glu	Arg	Lys	Asn	Asn	Asn	Lys	Arg	Trp	Tyr	Phe	Thr	Arg
1															

Glu	Gln	Leu	Glu	Asn	Ser	Pro	Ser	Arg	Arg	Phe	Gly	Val	Asp	Pro	Asp
20															

Lys	Glu	Leu	Ser	Tyr	Arg	Gln	Gln	Ala	Ala	Asn	Leu	Leu	Gln	Asp	Met
35															

Gly	Gln	Arg	Leu	Asn	Val	Ser	Gln	Leu	Thr	Ile	Asn	Thr	Ala	Ile	Val
50															

Tyr	Met	His	Arg	Phe	Tyr	Met	Ile	Gln	Ser	Phe	Thr	Arg	Phe	Pro	Gly
65															

Asn	Ser	Val	Ala	Pro	Ala	Ala	Leu	Phe	Leu	Ala	Ala	Lys	Val	Glu	Glu
85															

Gln Pro Lys Lys Leu Glu His Val Ile Lys Val Ala His Thr Cys Leu
100 105 110

His Pro Gln Glu Ser Leu Pro Asp Thr Arg Ser Glu Ala Tyr Leu Gln
115 120 125

Gln Val Gln Asp Leu Val Ile Leu Glu Ser Ile Ile Leu Gln Thr Leu
130 135 140

Gly Phe Glu Leu Thr Ile Asp His Pro His Thr His Val Val Lys Cys
145 150 155 160

Thr Gln Leu Val Arg Ala Ser Lys Asp Leu Ala Gln Thr Ser Tyr Phe
165 170 175

Met Ala Thr Asn Ser Leu His Leu Thr Thr Phe Ser Leu Gln Tyr Thr
180 185 190

Pro Pro Val Val Ala Cys Val Cys Ile His Leu Ala Cys Lys Trp Ser
195 200 205

Asn Trp Glu Ile Pro Val Ser Thr Asp Gly Lys His Trp Trp Glu Tyr
210 215 220

Val Asp Ala Thr Val Thr Leu Glu Leu Leu Asp Glu Leu Thr His Glu
225 230 235 240

Phe Leu Gln Ile Leu Glu Lys Thr Pro Asn Arg Leu Lys Arg Ile Trp
245 250 255

Asn Trp Arg Ala Cys Glu Ala Ala Lys Lys Thr Lys Ala Asp Asp Arg
260 265 270

Gly Thr Asp Glu Lys Thr Ser Glu Gln Thr Ile Leu Asn Met Ile Ser
275 280 285

Gln Ser Ser Ser Asp Thr Thr Ile Ala Gly Leu Met Ser Met Ser Thr
290 295 300

Ser Thr Thr Ser Ala Val Pro Ser Leu Pro Val Ser Glu Glu Ser Ser
305 310 315 320

Ser Asn Leu Thr Ser Val Glu Met Leu Pro Gly Lys Arg Trp Leu Ser
325 330 335

Ser Gln Pro Ser Phe Lys Leu Glu Pro Thr Gln Gly His Arg Thr Ser
340 345 350

Glu Asn Leu Ala Leu Thr Gly Val Asp His Ser Leu Pro Gln Asp Gly
355 360 365

Ser Asn Ala Phe Ile Ser Gln Lys Gln Asn Ser Lys Ser Val Pro Ser
370 375 380

DRAFT

Ala Lys Val Ser Leu Lys Glu Tyr Arg Ala Lys His Ala Glu Glu Leu
385 390 395 400

Ala Ala Gln Lys Arg Gln Leu Glu Asn Met Glu Ala Asn Val Lys Ser
405 410 415

Gln Tyr Ala Tyr Ala Ala Gln Asn Leu Leu Ser His His Asp Ser His
420 425 430

Ser Ser Val Ile Leu Lys Met Pro Ile Glu Gly Ser Glu Asn Pro Glu
435 440 445

Arg Pro Phe Leu Glu Lys Ala Asp Lys Thr Ala Leu Lys Met Arg Ile
450 455 460

Pro Val Ala Gly Gly Asp Lys Ala Ala Ser Ser Lys Pro Glu Glu Ile
465 470 475 480

Lys Met Arg Ile Lys Val His Ala Ala Ala Asp Lys His Asn Ser Val
485 490 495

Glu Asp Ser Val Thr Lys Ser Arg Glu His Lys Glu Glu Arg Lys Thr
500 505 510

His Pro Ser Asn His His His His Asn His His Ser His Lys His
515 520 525

Ser His Ser Gln Leu Pro Val Gly Thr Gly Asn Lys Arg Pro Gly Asp
530 535 540

Pro Lys His Ser Ser Gln Thr Ser Asn Leu Ala His Lys Thr Tyr Ser
545 550 555 560

Leu Ser Ser Ser Phe Ser Ser Ser Ser Thr Arg Lys Arg Gly Pro
565 570 575

Ser Glu Glu Thr Gly Gly Ala Val Phe Asp His Pro Ala Lys Ile Ala
580 585 590

Lys Ser Thr Lys Ser Ser Ser Leu Asn Phe Ser Phe Pro Ser Leu Pro
595 600 605

Thr Met Gly Gln Met Pro Gly His Ser Ser Asp Thr Ser Gly Leu Ser
610 615 620

Phe Ser Gln Pro Ser Cys Lys Thr Arg Val Pro His Ser Lys Leu Asp
625 630 635 640

Lys Gly Pro Thr Gly Ala Asn Gly His Asn Thr Thr Gln Thr Ile Asp
645 650 655

Tyr Gln Asp Thr Val Asn Met Leu His Ser Leu Leu Ser Ala Gln Gly
660 665 670

Val Gln Pro Thr Gln Pro Thr Ala Phe Glu Phe Val Arg Pro Tyr Ser
675 680 685
Asp Tyr Leu Asn Pro Arg Ser Gly Gly Ile Ser Ser Arg Ser Gly Asn
690 695 700
Thr Asp Lys Pro Arg Pro Pro Pro Leu Pro Ser Glu Pro Pro Pro Pro
705 710 715 720
Leu Pro Pro Leu Pro Lys
725

(2) INFORMATION FOR SEQ ID NO:51:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

TTCCCACCAA TGCTTTCC 18

(2) INFORMATION FOR SEQ ID NO:52:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

CCATCAGTTG ATACAGGGAT CT 22

(2) INFORMATION FOR SEQ ID NO:53:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

GGAATTCAAG AGGTTGTAAG ATGC 24

(2) INFORMATION FOR SEQ ID NO:54:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

ACACACAGAT GTGGTGAAAT GTACCCA

27

(2) INFORMATION FOR SEQ ID NO:55:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

GCATCTTACA ACCTTCTG

18

(2) INFORMATION FOR SEQ ID NO:56:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 28 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

GGAATTTCATG GAAAGCATTG GTGGGAAT

28

(2) INFORMATION FOR SEQ ID NO:57:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

CCTCCACTAC TGGTTTGCCT GG

22

(2) INFORMATION FOR SEQ ID NO:58:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 30 base pairs

- (B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

GGACTAGTAT AAATATGGCG TCGGGCCGTG

30

(2) INFORMATION FOR SEQ ID NO:59:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

GGAGATCTTA CATGTTCAATT CCTTGCGG

27

(2) INFORMATION FOR SEQ ID NO:60:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

GGAGACAAGT ATGTGCTACC TTGATGACA

29

(2) INFORMATION FOR SEQ ID NO:61:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

GGAATTGGGG CTGCTCCTCC ACTTTAG

27

(2) INFORMATION FOR SEQ ID NO:62:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

GGAATTCGCT GCTGGAGCCA CAGAA

25

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

GTGTCACTGA AAGAATAACCG

20

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

GGAATTCAGG TGGAGATAAA GCTGC

25

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

GCTCTAGATA AATATGGAGG GAGAGAGGAA

30

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

GGAATTCTTA CTTAGGAAGG GGTGGAAAGTG

30

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

GGAATTCTTA CTTAGGAAGG GGTGGAAAGTG GTGGAGGAGG TTAC

44

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

Ala Cys Ser Tyr Ser Pro Thr Ser Pro Ser Tyr Ser Pro Thr Ser Pro
1 5 10 15

Ser Tyr Ser Pro Thr Ser Pro Ser Lys Lys
20 25